FINAL REPORT

TO

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

Can Contingent Valuation Measure Passive-Use Values?

EPA Grant Number: R824688
Project Period: 10/1/95 – 12/31/98
Principal Investigator: William D. Schulze
Institution: Cornell University

by

Trudy Ann Cameron, Jeremy E. Clark, Robert G. Ethier, Gregory L. Poe, Daniel Rondeau, Steven K. Rose, and William D. Schulze

March 31, 1999

INDEX TO FINAL REPORT

Can Contingent Valuation Measure Passive-Use Values?

CHAPTER 1
Introduction

This report describes a project that attempts to compare and validate alternative elicitation methods for contingent valuation. The commodity used to test alternative methods was based on an actual green choice offering made by the Niagara Mohawk Power Company to fund non-renewable generation for 1200 homes and to plant 50,000 trees. Given the cooperation of the Niagara Mohawk Power Company, we were able to compare actual participation in the program solicited through a telephone interview with hypothetical participation as estimated from the various elicitation procedures employing both telephone and mail surveys.

The report is organized into six chapters where each chapter represents a stand-alone paper on a single aspect of the research that may be read independently by readers interested in the specific issues covered. Chapter 2 examines in general terms the demand revealing properties of one shot provision point mechanisms broadly similar to the one used by Niagara Mohawk Power Company using experimental economics methods. Chapter 3 presents an experimental test of the specific provision point mechanism utilized by Niagara Mohawk. We conclude, based on these two studies, that it is probable that a substantial share of demand is revealed by single shot provision points with large groups. Thus, comparing contingent values to the results of a provision point mechanism in a field setting is likely to be a superior methodology to that employed in the past, contingent values were compared to voluntary contributions. Chapter 4 compares the results of the actual phone solicitation with both contingent values obtained using dichotomous choice and open-ended willingness to pay from telephone interviews. Since we also employed mail surveys, and questions have been raised about the validity of mail surveys, we compare the results of the dichotomous choice telephone and mail surveys in Chapter 5. Finally, Chapter 6 compares all of the elicitation methods using both mail and telephone surveys with actual participation decisions. We present a brief summary of each of the remaining chapters below.

Chapter 2. Voluntary Revelation of the Demand for Public Goods Using a Provision Point Mechanism (Forthcoming in the *Journal of Public Economics*)

A one-shot provision point mechanism with money-back guarantee and proportional rebate of excess contributions is tested in an induced value framework and in experimental environments chosen to mimic field conditions. The results show that this relatively simple mechanism is empirically demand revealing in the aggregate when used with large groups of students who have heterogeneous valuations for the public good. Approximately demand revealing behavior was obtained under three alternative information conditions. These results are an important step in the design of a mechanism simple enough to allow field applications, but capable of efficiently providing public goods through voluntary contributions.

Chapter 3. The Private Provision of Public Goods -- Tests of a Provision Point Mechanism for Funding Green Power Programs, Cornell University Environmental and Research Economics Paper Series, 97-02, 1997.

This chapter utilizes field and laboratory experiments to test the use of a provision point mechanism to finance renewable energy programs, commonly known as green pricing programs. In contrast to most green pricing programs, relatively high participation is found in the field, while laboratory results suggest that demand revelation is achieved by the mechanism in a single shot environment with a large group of potential participants.

Chapter 4. Can Hypothetical Questions Predict Actual Participation in Public Programs? A Contingent Valuation Validity Test Using a Provision Point Mechanism (Submitted to *Journal of Environmental Economics and Management*)

Past field validity tests of contingent valuation have relied on voluntary contribution mechanisms to elicit actual willingness to pay, and are thus likely to overestimate hypothetical bias because of free riding in the actual contributions. This chapter argues that provision point mechanisms -- which have recently been shown to approximately reveal demand in large group public goods experiments -- should instead be used in contingent valuation validity testing, and employs such a mechanism in a validity study of green electricity pricing. Some upward hypothetical bias is found even when this improved mechanism is used. Calibration of hypothetical responses is also explored.

Chapter 5. A Comparison of Hypothetical Phone and Mail Contingent Valuation Responses for Green Pricing Electricity Programs (Forthcoming in *Land Economics*)

This study provides the first contingent valuation phone-mail comparison that meets current standards for response rates, draws from a general population, is relevant to the valuation of general environmental goods, and allows comparisons with actual participation rates. Social desirability effects are found to be more prevalent in phone responses to subjective questions, but do not appear to affect hypothetical participation decisions: calibrated and uncalibrated hypothetical participation rates are statistically similar across modes. As such, neither mode appears to dominate from the perspective of providing more valid estimates of actual participation decisions.

Chapter 6. Alternative Non-market Value Elicitation Methods -- Are the Underlying Preferences the Same? (Submitted to the *Journal of Environmental Economics and Management*)

In a unique survey, six different random sub-samples of respondents were presented with an opportunity to value the identical environmental good, each via a different elicitation method. The methods include: an actual purchase decision at a single price, a referendum format with differing prices, an open-ended format, a payment card format, a multiple-bounded format, and a stated choice among an extended set of five alternatives (including the basic good plus three additional alternatives). We employ a common underlying indirect utility function and a stochastic structure that is also fully compatible across methods, allowing us to pool all of these different types of valuation data in one unified model. The different types of valuation data are entirely compatible with the same underlying set of homogeneous preferences, providing heteroscedastic errors across elicitation methods are permitted.

CHAPTER 2
Voluntary Revelation of the Demand for Public Goods Using a Provision Point Mechanism* (Forthcoming, *Journal of Public Economics*)

1. Introduction

Benevolent organizations, clubs, associations and at times even governments and industry rely on citizen action and voluntary monetary contributions to provide a wide variety of public goods. Unfortunately, standard theoretical models of voluntary public goods provision make strong predictions of under-contribution by individuals. Although individuals do contribute voluntarily toward the provision of public goods, the evidence provided by two decades of experimental research clearly supports the prediction of under-contribution [Ledyard (1995), Davis and Holt (1993)]. Relying on funding mechanisms that inaccurately reflect contributors' preferences suggests that socially desirable public goods are produced at sub-optimal levels and underscores the need for a contribution mechanism capable of revealing the demand for public goods in natural conditions.

Such a mechanism has thus far eluded researchers. Some public goods mechanisms such as the Groves-Ledyard [Groves and Ledyard, (1977)] and Smith Auction [Smith, (1979 and 1980); Coursey and Smith, (1984); and Harstad and Marrese, (1982)] have been shown to induce optimal production of public goods in laboratory settings after a number of rounds of repeated play. Unfortunately, these are far too complex and impractical for implementation in field conditions where pragmatism dictates the use of a simply understood one-shot mechanism [Davis and Holt, (1993); Alston and Nowell (1996)]. An alternative mechanism, the Voluntary Contributions Mechanism (VCM) has the simplicity required for field applications but consistently produces contribution levels 40 to 60% below the optimum[1].

In this paper, we report the results of a series of laboratory experiments in which we

***Authors: Daniel Rondeau, William D. Schulze and Gregory L. Poe**

1 See Ledyard (1995) for a thorough survey of experimental research on the Voluntary Contributions Mechanism.

explore the performance of a one-shot provision point mechanism (PPM) with money back guarantee (MBG) and a proportional rebate of excess contributions (PR) applied in experimental conditions that attempt to replicate field circumstances. These include the use of large groups of non-economics students and incomplete information.

With this mechanism, a public good of pre-determined size is provided only if the sum of contributions equals or exceeds its cost (the provision point). If contributions fall short of costs, they are completely refunded (the money back guarantee) whereas if they exceed costs, the excess is returned to each contributor proportionally to the share of their individual contribution in the total amount contributed (the proportional rebate). As we indicate in the next section, the PPM has been shown to increase contributions but has fallen short of inducing demand revelation. Yet, because of the design of previous experiments, little can be said with certainty about the performance of this mechanism in environments resembling plausible field conditions where a single shot is required and a potentially large number of individuals with heterogenous values can benefit from the provision of a public good.

The central objective of our research is therefore to test the performance of the PPM with a MBG and a PR in experimental environments that mimic key features and constraints encountered in the field. The provision point framework enables us to induce well defined individual values for a discrete public good, thus providing a basis to measure how variations in the environment affect contributions. In particular, we investigate the impact of group size and incomplete information on the proportion of induced value revealed by subjects. Group size has been shown to have a positive effects on contributions with other public goods mechanisms [Isaac et al. (1994), Rose et al. (1997)] while limited evidence exists on the effect of incomplete information in PPM experiments [Marks and Croson (forthcoming)]. Our results show that, regardless of the information condition in which the mechanism is tested, large groups of non-business students given heterogenous values provide an aggregate amount of contributions approximately equal to their induced demand for the public good.

After a brief review of past findings, we present two sets of experiments in which we examine in turn the effects of group size and variations in information conditions on the proportion of induced demand revealed by participants. Concluding remarks follow.

2. The Provision Point Mechanism

In a typical provision point experiment, participants are part of a group of N individuals taking part in a number of decision rounds. At the beginning of a round, each person in the group is given an initial balance of money (denoted by the letter I) and must decide how much of this money to keep and how much to allocate to a group fund (B_i, where the subscript indexes the individual's contribution). If the mechanism includes a MBG (first tested in similar experiments by Isaac et al. 1989) and the sum of contributions is below the provision point (PP), contributions are fully refunded and individual earnings are equal to the initial balance. Alternatively, if the group sum of bids equals or exceeds the PP, the group fund yields a return and individual's earnings for the round are the sum of the initial balance minus her contribution, plus a personal return from the investment fund (the induced value, V_i), plus a payment according to the rebate rule implemented. The proportional rebate rule used in this paper was first proposed by Smith (1980) as part of the Smith public good auction and was recently studied by Marks and Croson (1998) in a PPM setting. Under this rule, contributions in excess of the PP are returned to individuals in proportion to the share of their personal contribution relative to the total received. For example, someone whose contribution amounts to 5% of total donations would receive a rebate equal to 5% of the amount of contributions in excess of the PP. Hence, individual i's earnings (E_i) are given algebraically by

$$E_i = \begin{cases} I & \text{if} \sum_{j=1}^{N} B_j < PP \\ I - B_i + V_i + \dfrac{B_i}{\sum_{j=1}^{N} B_j}\left(-PP + \sum_{j=1}^{N} B_j\right) & \text{if} \sum_{j=1}^{N} B_j \geq PP \end{cases}$$

In this context, a socially desirable public good is created by the experimenter if the sum of induced values is greater than the provision point. This public good is efficiently provided if subjects make aggregate contributions that equal or exceed the provision point. Therefore, aggregate demand revelation, if it can be achieved obtained, would guarantee that all desirable public goods will always be provided and that the mechanism is perfectly efficient regardless of the PP and whether or not it is know by subjects.

Isaac, Schmidtz and Walker (1989), Suleiman and Rapoport (1992) and Dawes et al. (1986) report that simply creating a threshold cost of provision had a significant positive impact on contributions compared to similar VCM treatments. This result is testimony to the power of a provision point since, in the absence of a MBG, failure to reach the provision threshold results in a loss of contributions by individuals. A MBG would therefore seem to be a desirable form of insurance against such losses. Significant increases in contributions have been reported by Isaac et al. (1989) in experiments where subjects were free to contribute any amount toward the provision of the public good and by Rapoport and Eshed-Levy (1989) in experiments where participants could contribute a fixed amount or noting at all. In contrast, Dawes et al. (1986) found no improvements in contribution rates after adding a MBG in a similar binary decision environment.

Adding a rebate rule is another form of insurance guaranteeing that contributions in excess of the PP will not be lost by the group. Marks and Croson (1998) investigated the effects of alternative rebate rules in PPM experiments with MBG. They report that implementing the PR rule or using excess contributions to increase the scope of the public good both improved contributions. Here, we chose to implement the PR rule because it has a positive effect on contributions and, compared to the increased scope rule, it fixes the cost and benefits of the public good, providing experimental control over the subjects' values.

The full information game theoretic predictions for the PPM with MBG was derived by Bagnoli and Lipman (1989) and is further discussed by Marks and Croson (1998) for the case where a PR rule is added to the mechanism. In this game, any combination of individually rational contributions summing exactly to the provision point is an efficient Nash outcome[1]. It should be clear that when the benefit-cost ratio of the public good is greater than one, aggregate demand revelation is not a Nash equilibrium since any individual who contributed a positive amount would choose to unilaterally reduce the amount of such contribution given the chance (as long as the PP is still met). Hence, the PPM is not theoretically incentive compatible. Whether or not it is demand revealing becomes is therefore a purely empirical question.

Bagnoli and McKee (1989) conducted full information provision point experiments with MBG using small groups of subjects who were given unequal values for the threshold public good.

---

[1] Assuming that players are only motivated by their own gains, a contribution is individually rational if it does not exceed the players' value for the public good. This game also has a large set of inefficient equilibria. An inefficient equilibria is any vector of individual contributions where

1) $\sum_{j=1}^{N} B_j < PP$; and 2) $\left(\text{PP} - \sum_{j=1}^{N} B_j\right) > (V_i - B_i)\ \forall\ i$. That is, the sum of contributions is below the provision point and no individual can unilaterally increase his or her contribution to make the group reach the provision point while maintaining individual rationality.

These subjects contributed on average 78.7% of their value in the first round of the game. In a simpler design, Cadsby and Maynes (fortcoming and 1998) used a PPM without MBG or rebate and gave equal endowments and valuations to all subjects. In the first round of separate experiments, economics and business students contributed 60.6% of their true demand, nurses revealed 85%, groups of male-only students contributed 40.2% and female-only students contributed 56%. Finally, in a study of alternative rebate rules, Marks and Croson (1998) obtained aggregate contributions averaging 63.7% of induced value in the first rounds of treatments with a MBG and PR.[3]

With contribution levels ranging from 40 to 85% of induced demand, variants of the PPM yield some improvements over the 40 to 60% of optimal contributions usually obtained with the VCM but it leaves a substantial gap for desirable projects to go unfunded. However, the laboratory environments constructed for earlier tests of the PPM differ from plausible field conditions in many respects. Hence, we seek to analyze the performance of the PPM with MBG and PR in experimental environments that more closely resemble field conditions.

The key environmental conditions that set our experiments apart from previous research are related to group size, repetition of play, subject pool used, distribution of benefits from the public good, and information available to participants. Whereas previous research was conducted with small groups of 5 to 12 subjects, we test the performance of the mechanism in groups of up to 50

[3] Rapoport and Suleiman (1993), Asch, Gigliotti and Polito (1993), Marks and Croson (forthcoming) and Croson and Marks (1996) have also conducted interesting provision point experiments. Unfortunately, individuals in these studies were given initial balances lower than their value for the public good. Hence, they faced an income constraint that did not allow them to reveal their true demand and precludes an unbiased analysis of demand revelation in these experiments. Similarly, Isaac, Schmidtz and Walker (1989) added a provision point and a money back guarantee to the VCM environment of Isaac, Walker and Thomas (1984). The mechanism used an extended benefit rule whereby contributions beyond the provision point increase the scope of, and hence the benefits from the public good. Additional benefits beyond the provision point eliminates the well defined demand against which to assess demand revelation performance.

subjects. In an equally important departure from prior research, we limit our experiments (with the exception of a control group) to a single round of decision making. This approach removes any possibility of strategic behavior that may exist in early rounds of repeated games. In the experiments reported in this paper, we assign subjects to one of several payoff conditions. Such heterogenous values are a characteristic of field conditions that allows us to induce a controlled, downward sloping demand curve for the public good. Of all experiments previously conducted, only Bagnoli and McKee's design combined heterogenous values with a sufficient money endowment for subjects to reveal their demand for the public good. Finally, for our second set of experiments, we recruited subjects with a more diverse background than the economics and business students traditionally used in experimental economics. This sets aside questions regarding the possible effects of economics training on behavior in these treatments. Of the research cited earlier, only Cadsby and Maynes have systematically chosen non-economics students in their treatment with nurses.

The resulting combination of environmental features differs substantially from any laboratory setting previously assembled for studying the PPM. This basic environment will be modified to test the effects of group size and incomplete information on contribution levels.

3. The Effect of Group-Size on Contributions

Isaac, Walker and Williams (1994) found that individuals in groups of 40 and 100 individuals contributed a significantly larger proportion of their endowment to a VCM public good than did subjects in groups of 4 and 10. The only other large group (n=100) experiment we are aware of was conducted as part of our own research program but in a different context. In Rose et al. (1997), we report that a PPM with MBG where subjects could only contribute a fixed amount (or not at all) produced aggregate results consistent with demand revelation: on average, individuals with values higher than the fixed amount contributed while those with values lower did not . However, because of the constrained contribution level, questions remain as to whether group size effects carry over to the PPM with continuous contributions, MBG and PR. We address this question first by comparing contribution levels in groups of six and fifty students.

Design. Four "pen and paper" experiments were conducted with lower division Cornell University students. The first three of these were small group control experiments while the fourth was carried out with a large number of subjects. For each of the small group experiments, six volunteer students who had never participated in economics experiments were recruited from an introductory economics class. It was emphasized in recruiting students that no knowledge of economics was required to participate.

At the beginning of the session, subjects read instructions describing the experiment and their task but were not given complete information about the parameters of the game.[4] They knew the size of the group and that all participants had an equal endowment of 500 experimental cents. They also knew their potential private payoff from the public good. However, they were told that this payoff had been randomly selected without being given any information about the distribution of values. They were only informed that other subjects may not have the same payoff. This feature mimics field conditions where individuals value public goods differently but are generally unaware of other people's values [Alston and Nowell (1996)].

The level of the provision point was not disclosed either. It was only announced that it had been randomly drawn from an unspecified distribution. Withholding information about the PP prevents subjects from making their contribution decision based on an "equal cost share" strategy

[4] Instructions are available from the authors.

(e.g. dividing the cost of the project equally among participants)[5] and may help raise contributions. Finally, participants in the small group experiments knew that the game would be repeated, but were not told how many times. For the purpose of this paper, however, we will only be interested in results from the first period since it is most comparable to the single shot large group experiment that follows[6].

Individual payoffs from the group fund, if the provision point was met or exceeded, were the randomly drawn numbers $2.12, $2.42, $3.69, $3.72, $3.76 and $3.90 experimental dollars, for total benefits (aggregate induced demand) from the public good of $19.63. The randomly drawn provision point was $7.53, creating a benefit-cost ratio of 2.6. Experimental earnings were exchanged at the rate of one dollar = $0.25 (US).

The large group experiment was conducted with fifty students from a different undergraduate economics class. Rather than recruiting on a voluntary basis, all students in the class participated in the experiment[7]. This experiment was limited to a single period of play. The one-shot game better

---

[5] Share calculations are rarely possible in the field, except for club goods. The information condition where the group size is known but the PP is unknown has real life parallels. An example comes from the Niagara Mohawk Power Company which recently offered its 1.2 million residential customers a green choice program in which the final cost of the project was to be determined through competitive bidding [Rose et al. (1997)].

[6] As in the one-shot application, subjects in the first period of a repeated game have no experience with the mechanism. However, repeated play may encourage signaling in early rounds, thereby increasing contributions in this treatment. If this were the case, the direction of the bias would make it more difficult to demonstrate that group size positively affects contribution levels. This approach is therefore conservative.

[7] All large group experiments reported in this paper have been conducted with entire classes of students. Conducting experiments with entire classes avoids the risk of self-selection bias inherent to the recruitment of volunteers. The experiments were conducted at the beginning of a regular class by a guest lecturer and his research assistants, none of whom had prior contacts with the students or would be involved in the student's grading. Participation was not mandatory and it was emphasized that the experiment was performed for research purposes only (although aggregate results would be reported at the end of the class) and that individual answers and earnings would remain strictly confidential. Finally, the regular instructor was not involved in conducting the experiment. We believe that these procedures reduce the possibility that expectational effects bias the data.

conforms to field conditions and allowed us to increase individual stakes fourfold by adjusting the exchange rate to one experimental dollar = $1 (US) (behavior should not be affected by such a monotonic transformation). The PP was scaled up to $62.75, an increase proportional to the change in the number of students in the group but this change cannot affect the results since the PP remained unknown to subjects. The experiment was otherwise identical to the small group control experiments. The same six randomly drawn induced values were used and the information position of subjects was unchanged.

Results. We use several indicators to report the performance of the mechanism. However, we are particularly interested in the mean and median of B_i / V_i , the proportion of individual induced value contributed to the group fund. We will also pay a special attention to the ratio of aggregate demand revealed to aggregate demand induced ($\sum B_i / \sum V_i$).

Pooling the data from the three small group experiments we find that individuals contributed on average 64% of their induced value, with a median of 71.8%. The ratio of aggregate revealed demand to total induced demand is 66.7%. This ratio falls within the range of 40.2% to 85% reported in previous research. Thus, we feel comfortable that our design and instructions are comparable to earlier PPM experiments and provide an adequate basis of comparison for the results of the large group treatment. A summary of the results is presented in Table 1.

Table 1
Summary data
Small and Large Group Comparison

Experiment ID	**Small Groups (Pooled Data)**	**Large Group**
Number of Subjects (N)	18	50
Mean V_i (cents)	327	325.8
Mean Contribution (cents) (SD)	218.1 * (128.2)	349.0 (135.5)
Median Contribution (cents)	200	380
Mean B_i/V_i (SD)	64.4% ** (32.1%)	110.0% *** (48.5%)
Median B_i/V_i	71.8%	104.0% ***
% of Demand Revealed	66.7%	107.1%

* different from 300 at the 5% significance level
** different from 100% at the 5% significance level
*** different from the small group result at the 5% significance level

In contrast, participants in the large group experiment approximately revealed their demand for the public good. The mean and median proportion of value contributed by individuals to the group fund were respectively 110% and 104%. The mean of 110% is not statistically different from 100% at the 5% significance level (t=1.459). On the other hand, we reject the hypothesis that the mean from the large group is equal to the mean of 64% of value obtained in the small group experiments (t=3.788)[8]. Parallel tests on the medians confirm theses results.[9] The medians of the small and large group experiments are significantly different from one another at the 5% significance level (Mann-Whitney z=3.753). A 95% confidence interval around the large group's median is bounded by 100.00% and 117.92%, clearly containing the value of 100% we would expect for a perfectly demand revealing mechanism. Hence, all tests support the conclusion that contribution levels in the large group experiment are different from those obtained from small groups, but not different from aggregate demand revelation. In the aggregate, the ratio of revealed to induced demand in the large group is 107%, a close approximation to demand revealing behavior.

Figures 1 and 2 illustrate the results. In these figures, the bars represent individual induced values and are graphed in descending order to form the induced demand for the public good. Individual contributions are also ordered from high to low and plotted as a line to represent the revealed demand curve. These graphs vividly illustrate the upward shift in the revealed demand curve obtained in the large group experiment. The overbidding and capping of contributions in the large group experiment is also visible in the upper left hand corner of Figure 2. These large bids roughly compensate for the cheap riding observable at the other end of the curve where induced demand is above revealed demand.

---

[8] The difference is significant at the 5% confidence level. The degrees of freedom for this test were adjusted to 47 to account for the statistical inequality of variances in the small and large group data.

[9] Tests on means are biased by the fact that roughly a quarter of all individual contributions in the large group experiment appear to have been constrained by the initial endowment of $5. This truncates the distribution of contributions, restraining both the mean and variance of the individual bid to value ratios. However, since the direction of the bias is to lower the difference between the means of the two experiments, we maintain that the difference would still hold in the absence of the endowment constraint. Tests on medians are not affected by the truncation of contributions.

Figure 1
Induced and Revealed Demand
Small Group Experiments Pooled

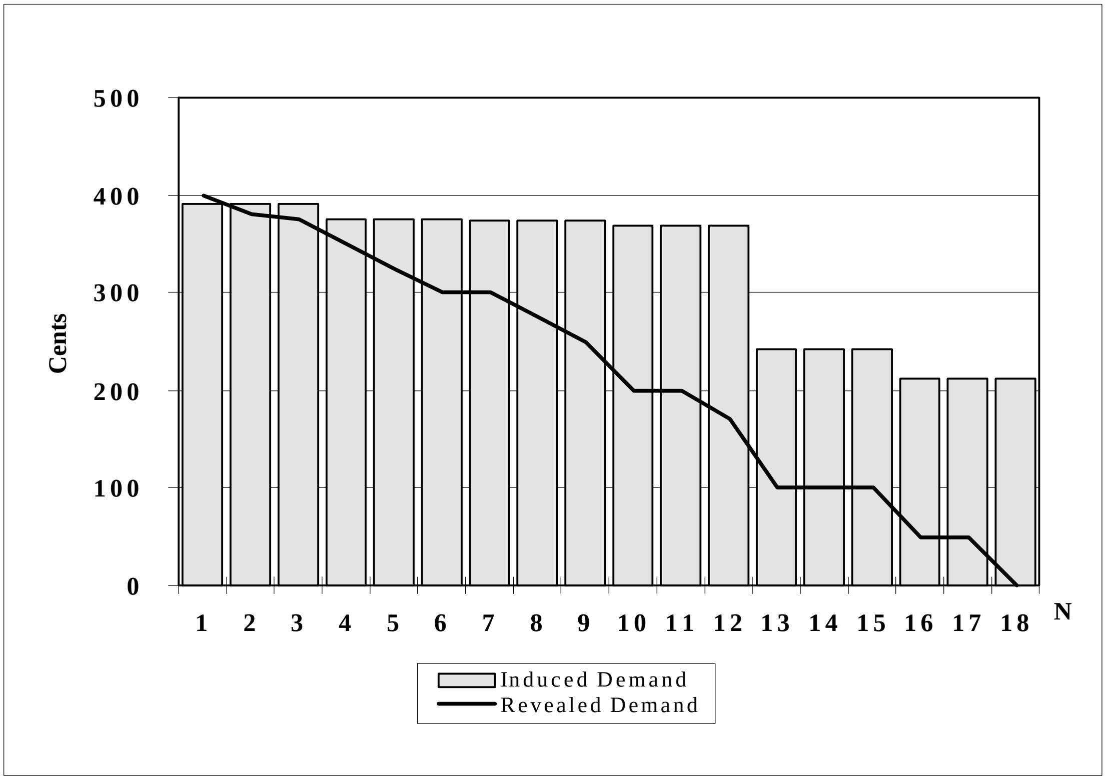
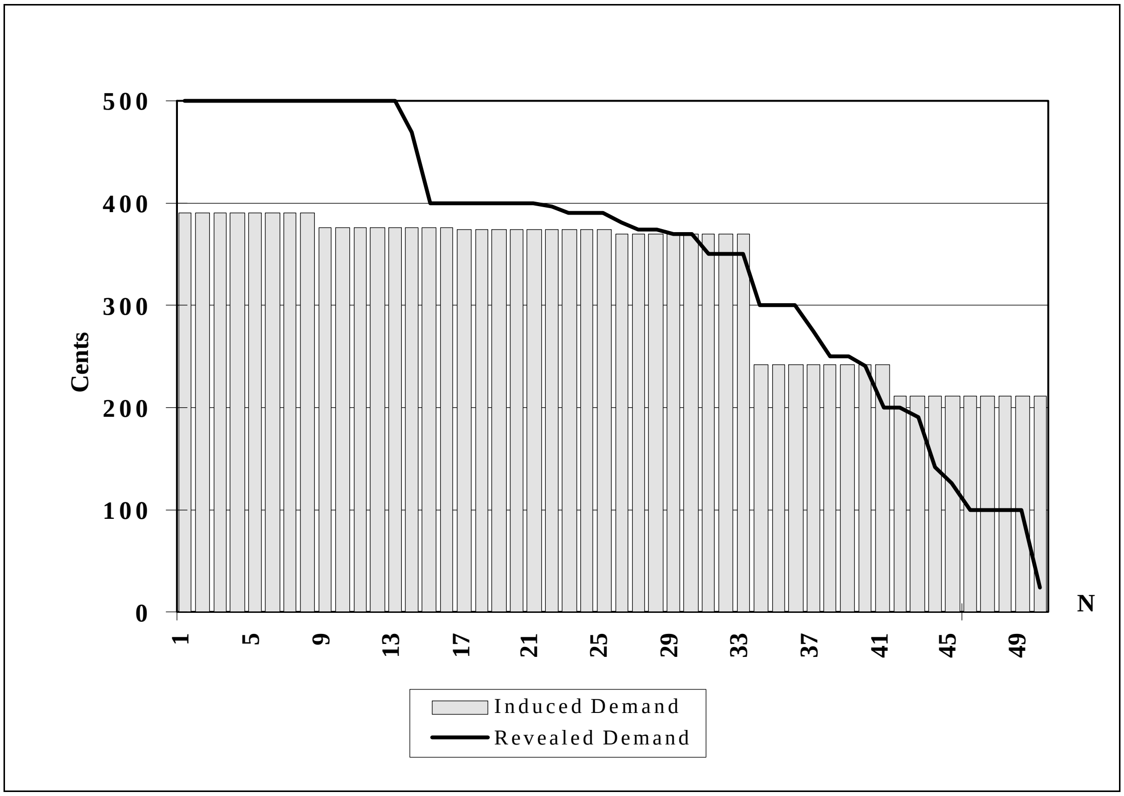


Figure 2
Induced and Revealed Demand
Large Group Experiment



While we attribute the increase in contributions primarily to group size, we must note that other design features did not remain constant between treatments and may be responsible for the observed differences. Subjects were from different classes, participants in the small group experiments were volunteers whereas the large group treatment was conducted with an entire class and the results reported for small groups are the donations in the first round of a repeated game as opposed to a single application of the mechanism. Notwithstanding these caveats, the data from these experiments demonstrate that it is possible to induce large groups to voluntary reveal their aggregate demand for public goods with a relatively simple mechanism. Next, we set out to replicate these results and test the effect of alternative information structures on subject behavior.

4. The Role of Incomplete Information in the Provision Point Mechanism

The objectives pursued with this set of experiments are 1) to replicate with non-economics students the results obtained in our first large group experiment and 2) to test whether alternative information conditions affect contribution levels. Specifically, we follow Bagnoli and Lipman and conjecture that when the number of subjects and the value of the PP are both known, subjects may show a greater tendency to choose a contribution level representing equal cost shares (PP/N, the symmetric Nash equilibrium). In a test of this conjecture, Bagnoli and McKee (1991) found empirical evidence of cost-sharing in full information experiments. Withholding information about the value of PP or N removes this focal point and may encourage demand revelation by forcing individuals to formulate a contribution strategy more strongly based on their private incentives rather than on the availability of a simple rule of thumb. This reasoning is similar to Bohm's (1972) argument that uncertainty about the cost of a public program puts "voters" in a situation in which incentives to using simple strategies leading to bias are absent.[10] Thus, we seek to explore the effect on contributions of removing information about the number of subjects and the level of the provision point.

Design. Three experiments were conducted with groups of 45 students enrolled in an introductory natural resources course. Approximately 5% of those students had previously taken an economics

---

[10] Marks and Croson (forthcoming) found that withholding information about other subject's values for the public good to prevent subjects from calculating the proportional cost-share $\left(PP * V_i \big/ \sum V_i\right)$ had no significant effect on the tendency of groups to adopt Nash behavior. This design still allowed subjects to calculate the equal cost share (PP/N) which is of interest to Bagnoli and Lipman, Bagnoli and McKee and to us.

class. Subjects were endowed with an initial balance of 600 cents and randomly assigned to one of five induced values ranging from $1.50 to $4.50, in increments of $0.75. The only difference between treatments was the information participants received about the number of subjects in their group and the level of the provision point. Subjects in group A were informed that 45 students in their group faced an investment cost of $45. Subjects in group B only knew the number of students in the group, and members of group C were only informed about the value of the provision point.[11] Therefore, this experiment not only tests the focal point hypothesis but also evaluates the performance of the PPM in alternative situations relevant to field applications.

Our original instructions were edited to accommodate these changes. Subjects in incomplete information treatments were told that the number withheld from them was "predetermined but unknown to you". Since the group size was unknown in one group, all instructions contained two examples of the proportional rebate rule. These examples used groups of size 2 and 200 respectively. Subjects were told that these were the minimum and maximum possible number of students in a group since the enrollment for the class was 200.[12] With n=45 and PP=$45, the focal point of $1 is easily computed by subjects in group A and was deliberately set low enough to make the cost sharing strategy rational for all subjects (all induced values were above $1).

[11] Each of these information conditions corresponds to a plausible public good situation. Group A corresponds to the funding of a club good with known cost. An example of such a situation occurred when the nordic ski club in Boulder CO raised money to maintain a bankrupt ski area for the winter. Members of group B represent a community of known size raising money for a project subject to cost uncertainty akin to the Niagara Mohawk Power Company program alluded to in Footnote 3. The information condition faced by members of group C are similar to many public allocation problems. A land trust soliciting donations from the public for the purchase and conservation of a tract of land falls in this category.

[12] Actual attendance for the class was 149. The 14 students who could not be accommodated in groups A, B, or C were put in a group D where both PP and N were unknown so that everyone in the classroom could participate in an experiment. We do not report the data from Group D since most students in this group were not present for opening remarks announcing the experiment and the general procedures that would be followed.

<u>*Results*</u>. Table 2 summarizes the descriptive statistics for each treatment. The first notable result is that there is no significant difference in behavior across alternative information conditions. The average proportion of individual value revealed by subjects ranges from 103% for group B to 132% for group C. Group A, which had the information to calculate equal cost shares falls between the two with an average of 110% of individual value revealed. Pair-wise means tests cannot detect significant differences between these values [$t_{A\text{ vs }B}$ = 0.381; $t_{A\text{ vs }C}$ = -0.970; $t_{B\text{ vs }C}$ = -1.392 (unequal variances)], and rank-sum tests comparing the median bid to value ratios of 86%, 93% and 100% also fail to indicate differences between the three treatments ($Z_{A\text{ vs }B}$ = 0.073; $Z_{A\text{ vs }C}$ = 0.892; $Z_{B\text{ vs }C}$ =1.00).

The number of $1 bids for groups A, B and C is respectively 7, 3 and 4. Hence, bids at the focal point are more frequent in group A. However, in tests of proportionality comparing group A to B and C, we cannot reject the null that the frequencies are equal ($p_{A\text{ vs }B}$ = 0.180; $p_{A\text{ vs }C}$ = 0.334). Based on this evidence, we conclude that the availability of information on N and PP did not create meaningful incentives to adopt a simple cost sharing strategy.

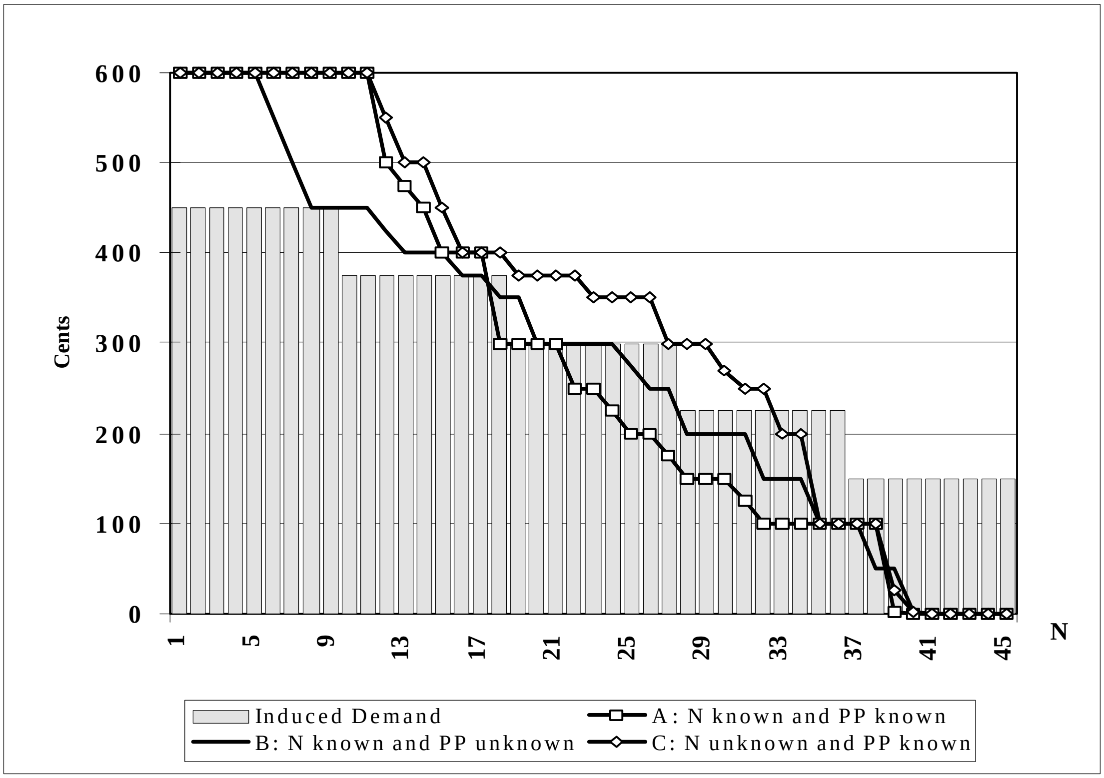
The most important finding of this paper is that, in large group situations, the ability of the provision point mechanism with MBG and PR to reveal aggregate demand appears robust to changes in experimental parameters, subject type and information provided to participants. The individual contribution to value ratios of 103%, 110% and 132% found for groups A, B and C are not different from 100% at the 5% significance level. Similarly, the medians, at 86%, 93% and 100% of induced value each generates a 95% confidence interval that includes 100%. We also note that the mean bids of $2.89 (t=0.332), $2.86 (t=0.494) and $3.38 (t=1.208) do not statistically differ from the mean induced value of $3.00. Finally, 95% confidence intervals around each of the median contributions of $2.50, $3.00 and $3.50 include the value of $3 consistent with a mechanism that, overall, produces demand revelation.

Table 2
Summary Data
Comparison Across Information Conditions

Experiment ID	A	B	C
Information Provided	(N) yes; (PP) yes	(N) yes; (PP) no	(N) no; (PP) yes
Number of Subjects (N)	45	45	45
Mean V_i (cents)	300	300	300
Mean Contribution (cents) **(SD)**	288.9 (221.2)	285.6 (190.3)	337.7 (210.8)
Median Contribution (cents)	250	300	350
Mean B_i/V_i **(SD)**	110.5% (99.2%)	103.2% (83.5%)	132.2% (112.0%)
Median B_i/V_i	86.1.%	93.3%	100.0%
% of Demand Revealed	96.3%	95.2%	112.6%

The total demand revealed by the three groups are 96%, 95% and 112% of total induced value. These ratios are comparable to the ratio of 107% found in our first large group experiment. Hence, our initial large group results were replicated in a total of three different information conditions, with a modified vector of induced values, and using non-business students. Figure 3 illustrates the results. As we previously reported, overbidding by some subjects essentially offsets the cheap-riding of others in all treatments. While the slopes of the revealed demand curves are poor indicators of true demand, the means and medians are accurate and can be used to infer the aggregate benefits of the public good. Since a number of subjects still appear to have been constrained by their endowment, median statistics should be given more weight than means.

Figure 3
Induced and Revealed Demand
Effects of Information Experiments



5. Conclusion

Using large groups in an induced value framework, we have shown that the provision point mechanism with money-back guarantee and proportional rebate of excess contributions can closely approximate demand revelation. The mean and median statistics for both, the absolute individual contributions and the proportion of induced value revealed were all statistically consistent with demand revealing behavior. To our knowledge, it is the first time that a simple one-shot public goods mechanism that allows a wide range of donation levels has elicited voluntary contributions approximately equal to the true value of a public good.

Overall, the cheap-riding of some subjects was compensated by the over-contributions of others in all four large group experiments presented in this paper. The fact that some subjects contributed amounts above their induced value appears to be irrational and suggests that additional research designed to test individual motives is required. It is possible that subjects have altruistic motives that cannot be directly controlled for in the laboratory. Some individuals may also erroneously interpret the MBG and PR as providing an insurance that earnings cannot be less than their initial endowment or that by contributing a large amount, it is possible to capture a larger part of contributions in excess of the PP. Palfrey and Prisbrey (1997) have recently suggested that, indeed, high contribution levels in early rounds of public goods (VCM) experiments can in part be explained by subject errors, but they also present evidence of the existence of a positive "warm glow" associated with the simple act of contributing toward the provision of a public good.

The provision point mechanism with money-back guarantee and proportional rebate is simple enough to provide hope that similar results can be replicated under field conditions. Yet, the failure of the same mechanism to reveal demand in prior research and in our own small group trials suggest that the key to fully understand the PPM resides not so much in the mechanism itself as in the environment in which it is applied. Some of the findings of this research point to group size as a determinant factor affecting contributions. Nevertheless, several additional experiments will be required before we can adequately understand how group size and other factors such as subject background, recruitment or self-selection, altruism or warm-glow affect the performance of the PPM, and assess its capacity to efficiently provide public goods in real world situations.

References

Alston, R.M., Nowell, C., 1996. Implementing the voluntary contribution game: a field experiment. Journal of Economic Behavior & Organization 31, 357-368.

Asch, P., Gigliotti, G.A., Polito, J.A., 1993. Free riding with discrete and continuous public goods: some experimental evidence. Public Choice 77, 293-305.

Bagnoli, M., Lipman, B., 1989. Provision of public goods: fully implementing the core through voluntary contributions. Review of Economic Studies 56, 583-601.

Bagnoli, M., McKee, M., 1991. Voluntary contributions games: efficient private provision of public goods. Economic Inquiry 29, 351-366.

Bohm, P., 1972. Estimating demand for public goods: an experiment. European Economic Review 3, 111-130.

Cadsby, C.B., Maynes, E., forthcoming. Choosing between a socially efficient and a free-riding equilibrium: nurses versus economics and business students. Journal of Economic Behavior and Organization.

Cadsby, C.B., Maynes, E., 1998. Gender and free riding in a threshold public goods game: experimental evidence. Journal of Economic Behavior and Organization 34, 603-620.

Coursey D., Smith, V.L., 1984. Experimental tests of an allocation mechanism for private, public or externality goods. Scandinavian Journal of Economics 86, 468-484.

Croson, R.T.A., Marks, M., 1996. Equilibrium selection: preplay communication and learning? Working Paper, Risk Management and Decision Processes Center, Wharton School of Business, University of Pennsylvania.

Davis, D.D., Holt, C.A., 1993. Experimental Economics. Princeton University Press, Princeton, NJ.

Dawes, R., Orbell, J., Simmons, R., van de Kragt, A., 1986. Organizing groups for collective action. American Political Science Review 8, 1171-85.

Groves, T., Ledyard, J., 1977. Optimal allocation of public goods: a solution to the 'free rider' problem. Econometrica 45, 783-809.

Harstad, R., Marrese, M., 1982, Behavioral explanations of efficient public good allocations. Journal of Public Economics 19, 367-383.

Isaac, R.M., Schmidtz, D., Walker, J., 1989. The assurance problem in laboratory markets. Public Choice 62, 217-236.

Isaac, R.M., Walker, J., Thomas, S., 1984. Divergent evidence on free riding: an experimental examination of possible explanations. Public Choice 43, 113-149.

Isaac, R.M., Walker, J., Williams, A., 1994. Group size and the voluntary provision of public goods: experimental evidence using very large groups. Journal of Public Economics 54, 1-36.

Ledyard, J.O., 1995. Public Goods: a survey of experimental research. In: Kagel, J.H., Roth, A.E. (Eds.), Handbook of Experimental Economics. Princeton University Press, Princeton, NJ. pp. 111-194.

Marks, M.B., Croson, R., 1998. Alternative rebate rules in the provision of a threshold public good: an experimental investigation. Journal of Public Economics 67, 195-220.

Marks, M.B. Croson, R., forthcoming. The effect of incomplete information in a threshold public goods experiment. Public Choice.

Palfrey, T.R., Prisbrey, J.E., 1997. Anomalous behavior in public goods experiments: how much and why? American Economic Review 87, 829-846.

Rapoport, A., Eshed-Levy, D., 1989. Provision of step-level public goods: effects of greed and fear of being gypped. Organizational Behavior and Human Decision Processes 44, 325-344.

Rapoport, A., Suleiman, R., 1993. Incremental contribution in step-level public goods games with asymmetric players. Organizational Behavior and Human Decision Processes 55, 171-194.

Rose, S., Clark, J., Poe, G.L., Rondeau, D., Schulze, W.D., 1997. The private provision of public goods: tests of a provision point mechanism for funding green power programs. Environmental and Resource Economics Working Paper, no. 97-02. Cornell University.

Smith, V.L., 1980. Experiments with a decentralized mechanism for public goods decision. American Economic Review 70, 584-599.

Smith, V.L., 1979. An experimental comparison of three public good decision mechanisms. Scandinavian Journal of Economics 81, 198-215.

Snedecor, G.W., Cochran, W.G., 1989. Statistical Methods - eight edition. Iowa State University Press, Ames, IA.

Suleiman, R., Rapoport, A., 1992. Provision of step-level public goods with continuous contribution. Journal of Behavioral Decision Making 5,133-153.

CHAPTER 3
The Private Provision of Public Goods: Tests of a Provision Point Mechanism for funding Green Power Programs*

1. Introduction

Despite market research that has uniformly predicted substantial customer interest in paying higher electric power rates to support renewable energy generation and environmental programs, experience with green pricing indicates that participation levels have not exceeded 1 to 2 percent (Byrnes *et al.*, 1995; Farhar and Houston, 1996).[3] Three explanations for this discrepancy seem possible. First, hypothetical market research studies of program support may have been upwardly biased. Second, most utility customers may have been unaware of such programs, in spite of attempts by electric utilities to inform them using bill inserts, mailed brochures and advertising. Note that market research, by necessarily informing customers of a potential green pricing program, inherently creates perfect awareness concerning the program in the sample population. As a result, forecasts derived from market research depend critically on assumptions about customer awareness which in turn depend on the effectiveness of marketing. A third possibility is that actual customer participation in green programs may have been lowered by free-riding, because participation has commonly been structured as a charitable voluntary contribution.

From the viewpoint of economics, the possibility of free riding in actual participation is of primary concern. Provision point mechanisms have been shown to have desirable theoretical properties (Bagnoli and Lipman, 1989) and to substantially reduce free riding in experimental tests when compared to the voluntary contribution mechanism (VCM) (Isaac, Schmidtz, and Walker, 1989; Suleiman and Rapoport, 1992; Dawes, Orbell, Simmons, and van de Kragt, 1986). There are also anecdotal reports of provision points being used to successfully resolve actual free riding problems (Bagnoli and McKee, 1991). Motivated in part by this literature, as well as by

***Authors: Steven K. Rose, Jeremy Clark, Gregory L. Poe, Daniel Rondeau, William D. Schulze**

recent utility industry interest in voluntarily funded green power programs, this paper reports the results of a paired field and laboratory experimental application of a provision point mechanism using a green pricing program implemented by Niagara Mohawk Power Corporation. Both theoretical and experimental economists have long hoped for a practical mechanism for the private funding of public goods (see for example Groves and Ledyard, 1977; Smith, 1980). This research is designed to test whether this goal can be realized given our current understanding of public good mechanisms.

In Section 2 we provide the specifics of the Niagara Mohawk Power Corporation *GreenChoice*™ program and the provision point mechanism used. The third section describes the field experiment and estimates a random utility model of actual program participation on the basis of individual characteristics. The primary advantage of the field experiment is that, by phoning customers, describing the *GreenChoice*™ program, and allowing them to sign-up or decline the offering on the phone, complete awareness is assured in the sample population. In spite of this control, it is still uncertain whether the sign-up rates observed in the field experiment (which are much higher than those of previous programs using voluntary contributions) reflect actual demand or if free-riding problems remain. Thus, in Section 4, we replicate the Niagara Mohawk Power Corporation mechanism in an induced value laboratory experiment under the assumption that, if the mechanism fails to eliminate free riding in the laboratory, then it will fail to eliminate free riding in the field. The hypothesis that this provision point mechanism eliminates free riding and induces demand-revealing behavior is tested by comparing individual and group contributions relative to induced values.[1] A random utility model is used to predict the probability of participation, but now as a function of induced value. Finally, Section 5 presents our conclusions concerning use of provision points for the private provision of public goods and discusses remaining issues.

2. The Niagara Mohawk Power Corporation GreenChoice™ Program

The Niagara Mohawk Power Corporation (NMPC), a public utility in New York State, sought to accelerate the development of renewable energy sources of electricity by offering its customers "green rates" as proposed by Moskovitz (1992, 1993). Moskovitz argued that customers would voluntarily sign up and agree to pay higher electricity rates if the additional money collected were earmarked to support renewable energy projects or other environmental activities. Economists were quick to point out that the selection of such a rate by a customer would be a charitable contribution since the mechanism proposed by Moskovitz would allow free riding (see Schulze, 1994).[2] NMPC in turn developed the *GreenChoice*™ program, using a modified contribution mechanism in an attempt to reduce free-riding incentives.

The mechanism adopted by NMPC employed three features that have been tested in the experimental literature. First, it contained a *provision point* of $864,000 to be raised through customer contributions. This minimum level of funding would provide for the construction of a renewable energy facility to serve 1,200 homes, and for the planting of 50,000 trees in the NMPC service area. The addition of a provision point to a voluntary contribution mechanism adds multiple, efficient Nash equilibria at the threshold, and has been shown to increase individual pledges towards the provision of public goods. Unfortunately, if the threshold is not met, a provision point results in a complete loss of efficiency, unlike the VCM (Isaac, Schmidtz and Walker, 1989).

Second, NMPC's funding mechanism offered a *money back guarantee* to customers which assured them that, if contributions failed to reach the threshold, all money collected would be refunded. The money-back guarantee provided insurance to potential contributors against the risk of losing their contributions should the provision point not be met. In experiments where subjects can contribute all or none of their endowment to a public good, Dawes et al. (1986) find no evidence to support the use of a money-back guarantee. However, in an environment where subjects can contribute any amount, Isaac, Schmidtz, and Walker (1989) report that the guarantee significantly increases contributions.

Third, the mechanism offered the possibility of *extended benefits*. Money collected in excess of the provision point would be used to extend benefits, or increase the production of the public good. Here, excess contributions were to be used to increase the number of homes served with renewable energy or to plant more trees. Extending benefits beyond the provision point does not modify individual incentives in theory, but simply creates a VCM environment beyond the threshold (Marks and Croson, 1996). Marks and Croson refers to this use of excess contributions as a "utilization rebate" rule. In evaluating alternative rebate rules for provision point mechanisms experimentally, Marks and Croson finds that offering extended benefits has the greatest positive effect upon group contributions.

One theoretically undesirable feature of NMPC's mechanism was that, to legally qualify as a rate offering, the program could only be offered at a posted price. Thus, customers could choose only to contribute a fixed amount of $6.00 per month or not participate at all. A posted price is undesirable because it does not allow households to self-select a monthly fee that better represents their preferences for the program. Note that, despite the posted price, the mechanism does not reduce to a referendum, because the only individuals to pay are those who choose to participate.

Interestingly, the only other green pricing programs to use a provision point mechanism of which we are aware were fully subscribed. Traverse City Light and Power attempted and completed a windmill project using a funding mechanism similar to NMPC's, except that it did not offer extended benefits. Participation was instead curtailed after the program's provision point was successfully reached with 200 customers at an estimated residential premium of $7.58 per month (23 percent of the average residential bill) (Holt and Associates, 1996a). The City of Fort Collins also used a series of provision points to solicit funds for up to three separate wind turbines. (Holt and Associates, 1996b). To date, enough customers have agreed to pay an estimated average premium of $10 per month to exceed the minimum provision point established to fund two turbines (Clements-Grote, 1997; Holt and Associates, 1997).

In comparing these offerings with the *GreenChoice*™ program it is important to note that there are substantial differences in magnitude and scope. Both the Fort Collins and Traverse City programs were small, locally based programs able to focus on well-defined projects, so that awareness was easily achieved. In contrast, the NMPC program, although initially intended to be offered only in the Buffalo area, had to be offered, for legal reasons, to NMPC's entire service area, which covers well over fifty percent of the area of New York State. Thus, marketing became a major impediment to the program.

Unfortunately, though the *GreenChoice*™ program was formally approved by the New York Public Service Commission, it was ultimately suspended before completion because NMPC developed serious financial difficulties and was unable to promote customer awareness of the program.   Before suspension, the program was briefly mentioned in a bill insert and described in a brochure sent to about three percent of NMPC's customers. Most of the planned marketing campaign, including a substantial advertising budget and tree plantings at public schools throughout the service territory, was canceled. Before the program was terminated, however, we were able to conduct a field experiment with NMPC customers.

3. Field Experiment

3.1. Experimental Design

The field experiment was conducted as part of a larger National Science Foundation/Environmental Protection Agency research effort to investigate environmental values for public programs (Poe, Clark, and Schulze, 1997). A telephone survey was utilized to attempt to contact a random sample of 206 households in the Buffalo area.[3] The telephone survey began by screening customers to identify the person in the household who usually pays the NMPC electric bill. Once that person is on the phone, the interviewer describes the purpose of the survey and the sponsors of the study. The individual is then asked to rate NMPC's service. This allows the small number of dissatisfied customers to vent frustration before answering the remaining questions. Customer awareness of the *GreenChoice*™ program is obtained next, and then the goals of the program are described in turn. As the goals are described, the respondent is asked:

> *How interested are you in the goal of replacing fossil energy with renewable energy sources? On a scale from 1 to 10, where 1 is not at all interested and 10 is very interested, how interested are you?*

and later:

How interested are you in the goal of planting trees on public lands in upstate New York? As before on a scale from 1 to 10, where 1 is not at all interested and 10 is very interested, how interested are you?

The funding plan is then described as follows:

The GreenChoice program would be funded voluntarily. Customers who decide to join the program would pay an additional fixed fee of $6 per month on their NMPC bill. This fee would not be tax deductible. Customers would sign up or cancel at any time. While customers sign up, NMPC would ask for bids on renewable energy projects. Enough customers would have to become GreenChoice partners to pay for the program. For example if 12,000 customers joined the first year, they would invest $864,000, which would allow Niagara Mohawk to plant 50,000 trees and fund a landfill gas project. The gas project could replace all fossil fuel electricity in 1,200 homes. However, if after one year, participation were insufficient to fund GreenChoice activities, Niagara Mohawk would cancel the program and refund all the money that was collected.

The program description was taken more or less directly from the program brochure prepared by NMPC. Note that NMPC was deliberately vague about the exact level of the provision because the renewable energy project was to be sent for competitive bid.

The survey then asks respondents whether the features of the funding program make them more or less interested in the program (see section 3.2 for details). This is followed by the participation question. It is phrased as follows:

You may need a moment to consider the next couple of questions. Given your household's income and expenses, I'd like you to think about whether or not you would be interested in the GreenChoice program. If you decide to sign up, we will send your name to Niagara Mohawk, and get you enrolled in the program. All your other answers to this survey will remain confidential. Does your household want to sign up for the program at a cost of $6.00 per month?

Although actual monies were never collected because the program was suspended, this sign up now/pay later approach corresponds with the following stepwise process typically used in green pricing programs: 1) potential projects are described; 2) subscriptions from customers are elicited through direct marketing, bill inserts and advertising; and 3) money is collected through regular billing. Experience from the Traverse City project suggests that the payment to intention ratio is very high--in that case, Traverse City Light and Power found that approximately 5% of those who originally signed-up reneged.

The survey ends with socioeconomic questions useful for modeling demand.

3.2. Results and Analysis

Of the sample of 206 households, contact was made with 179.[4] Of these, 34 refused to participate and three could not complete the questionnaire. Thus, 142 respondents completed the survey, yielding a response rate of 69% of the base sample. Of the 142, 29 signed up for the program, resulting in a participation rate of 20.5 percent. If we assume that the 37 households who refused or could not complete the survey would also have refused the program, the participation rate would fall to 16.5 percent. Both these estimates stand in marked contrast to the actual sign-up rate of less than 0.1 percent observed by NMPC throughout the period *GreenChoice*™ was offered. As discussed previously, this low participation was likely caused by the minimal marketing and low customer awareness of the program. Indeed, none of the 142 randomly sampled respondents in our survey recalled having heard about the program. Participation rates of 16.5 and 20.5 percent are consistent with a preliminary market evaluation of the NMPC service area conducted by the Research Triangle Institute (RTI) (Wood *et al.* 1994), which estimated that with full awareness there was a 17 percent probability of adopting a green planting program at a $6 monthly premium. The RTI data were taken from a sample that over sampled "green" customers, since such customers were regarded as the target group for an actual program. Based on prior information, approximately 25 percent of urban NMPC customers were classified as “green”.

It is important to note that a participation rate of 16%-20% is, however, substantially higher than the 1% needed to fund *GreenChoice*™ (12,000 of a total of 1.2 million NMPC customers), and those observed in the majority of actual green pricing experiments reported in the literature (Baugh *et al.* 1995; Brynes *et al.* 1995; Holt and Associates, 1996; Farhar and Houston 1996). As suggested earlier, however, there are notable differences between our experiment and the majority of previous studies. First, reported participation rate estimates have not generally been adjusted to account for program awareness, which was controlled in our study at 100 percent. Instead, participation rates have typically been defined over total customer base or over the base of customers targeted with direct mailings. Previous participation experiments have also (with the two exceptions noted previously) relied on voluntary contributions rather than the provision point mechanism used here.

To investigate individual specific factors associated with participation decisions, the linear logistic distribution, which can be derived from a random utility model (McFadden, 1976), is assumed to characterize individual decisions,

$$\text{(1)} \qquad \Pr\{\text{“Yes” response}\} = \frac{1}{1 + e^{-\underline{\alpha X}}}$$

where $\underline{X}$ depicts a vector of covariates characterizing individuals and their perceptions of the program (including a constant term), and $\underline{\alpha}$ is the corresponding set of coefficients to be estimated.

Assuming this logistic distribution, participation decisions are modeled as a function of three categories of covariates elicited in the questionnaire. The first concerns respondents' perceptions of the program's worth. Respondents registered their interest in the twin goals of the *GreenChoice*™ program -- replacing fossil fuels and planting trees in upstate New York -- using a scale of one (*"not at all interested"*) to 10 (*"very interested"*) for each goal.[5] It is expected that the sign on these variable would be positively correlated with the probability of joining the program.

The second category of covariates includes variables specific to the respondent, such as sex (Male=1), age (Years), education (College Graduate or higher =1), and recent financial support of environmental groups (Yes=1). Such characteristics are widely used as explanatory covariates in the environmental valuation literature. Based on this literature, it is expected that age will be negatively correlated with WTP while recent financial support for environmental groups would be positively correlated with joining the program. The other variables have provided mixed results in the literature. As noted earlier, individual perceptions of NMPC service were elicited using a one (*"unfavorable"*) to 10 (*"very favorable"*) scale and included as a covariate in this analysis.

The final category of covariates concerns respondents' perceptions of the provision point mechanism itself. After hearing of the funding provision point and money back guarantee, respondents were asked the following two questions:

> *Does the fact that a minimum level of customer participation is required for GreenChoice to operate make the program of less interest to you, more interest, or does it not affect your interest?*
>
> *Does the fact that Niagara Mohawk would refund all the money it collects -- if support is insufficient -- make GreenChoice of less interest to you, more interest, or does it not affect your interest in the program?*

These variables are admittedly *ad hoc*, in the sense they do not proxy for the value of the program. However, they do provide information about perceptions regarding specific components of the provision point mechanism. Over 55 percent responded that their interest was not affected by including a provision point and about 16 and 27 percent indicated that it increased or decreased their interest in the program, respectively. In contrast, the money back guarantee was widely favored: only 9 percent of respondents indicated that this attribute reduced their interest in the program, while 46 percent indicated that it increased their interest. For the purpose of modeling the participation decision, these response categories were re-coded as binary variables assigned '1' if the *"more interest"* option was selected, and zero otherwise. We expect their estimated coefficients to be positive.

The logit model of program participation is reported in Table 1, together with the sample means, standard deviations, and the expected signs of the estimated coefficients of all the explanatory variables described above. Given the single \$6 threshold, the estimation results are fairly strong: 80 percent of the responses are correctly predicted and the overall likelihood greatly exceeds the critical value (LR=31.03 > 14.68 = $\chi^2_{0.10}(9)$).

Considered jointly, the estimated coefficients on the two program goals are significant using a likelihood ratio test (LR = 7.23 > 4.61 = $\chi^2_{0.10}(2)$), leading to the conclusion that there is a positive response to the tree-planting and renewable energy objectives of the NMPC program. Comparison of the individual coefficient estimates suggests that, in spite of the observation that more people favored the tree planting objective, interest in fossil fuel replacement is a more significant predictor of participation decisions. The implication is that tree programs will have road based general support, but that interest in the fossil fuel component will be the significant

Table 1. Estimated Logit Models of NMPC Phone Participants

Variable [Scale]	Mean	Expected Sign	Estimated Coefficients
Constant	1	n.a.	-4.386 (2.184)**
Replace Fossil Fuel [1-10]	6.27 (2.82)	+	0.233 (0.118)**
Plant Trees [1-10]	8.35 (2.18)	+	0.216 (0.186)
Sex [Male = 1]	0.46 (0.50)	?	0.954 (0.517)*
Age [Numeric]	55.09 (15.70)	-	-0.0396 (0.0192)**
Give to Environment [Yes = 1]	0.19 (0.39)	+	0.666 (0.624)
College Graduate [Grad = 1]	0.45 (0.50)	+?	0.002 (0.546)
Rating of NMPC Service [10=very good]	8.49 (1.67)	+?	0.082 (0.644)
Min. Participation [More Interested = 1]	0.17 (0.38)	+	1.416 (0.588)**
Money Back Guarantee [More Interested = 1]	0.47 (0.50)	+	-0.098 (0.550)
n	128		128
Likelihood Ratio χ^2			31.03***
Percent Correctly Predicted			80

Numbers in () are standard errors.

*,**, and *** indicate significance levels of 10, 5, and 1 percent, respectively.

explanatory factor in participation decisions. This finding is consistent with the NMPC market research (Wood *et al.*, 1994).

A joint test of the null hypothesis that restricts all demographic coefficients to zero was rejected at the 10 percent level (LR = 10.28 > 9.24 = $\chi^2_{0.10}(5)$). The estimated coefficients on respondent attributes vary in significance, consistent with other studies in the environmental valuation literature. Age was negatively correlated with participation, a factor that may be attributed to the life cycle hypothesis of value in which potential use values decline with age (Cropper and Sussman, 1990). This negative relation may also be associated with the fact that age is also inversely correlated with income in this data set.[6] The finding that male respondents had a higher likelihood of participation contrasts with evidence suggesting that this variable is not substantially related to environmental concerns (Van Liere and Dunlap, 1980). The coefficients on the other socio-demographic covariates were not significantly different from zero.

From our perspective, the coefficients on the funding mechanism variables are of considerable interest, despite their *ad hoc* nature. Considered jointly, these variables are significant ((LR = 5.84 > 4.61 = $\chi^2_{0.10}(2)$). In particular, interest in the provision point mechanism is a significant, and positive, explanatory variable in participation decision. The minority of respondents with interest in that feature clearly had a higher participation rate, suggesting that addition of this feature increases the likelihood of funding. In contrast, interest in the money back guarantee is not a significant explanatory variable in the estimated model in spite of the fact that there appears to be a widespread interest in the money back guarantee.

In summary, modeling of participation decisions indicates that the content and structural attributes of the NMPC mechanism are influential in participation decisions. The program goals of replacing fossil fuel energy and planting tree are important to participation decisions, particularly the former. In addition, the provision point feature increases participation.

4. Laboratory Experiment

4.1. Experimental Design

The provision point mechanism adopted by NMPC appears, given the field experiment results, to yield a high participation rate with full consumer awareness. In addition, there seems to be a consistent relationship between individuals' stated preferences and program involvement. Nevertheless, without direct knowledge of individual valuations, we have no way of knowing how successful the mechanism is in eliminating free riding or if the mechanism is demand revealing. A laboratory experiment was thus designed to test this funding mechanism in an environment where program values could be induced. If this mechanism fails to eliminate free riding in the laboratory, then we would expect it to fail to eliminate free riding in the field. Note that provision point mechanisms theoretically have Nash equilibria where costs are just covered by contributions. Often, in laboratory experiments with small groups, subjects just miss the provision point by slight under-contribution, a behavior termed “cheap riding” (Bagnoli and Lipman, 1989). In contrast, there is some evidence that large groups reveal demand when faced with a single shot provision point mechanism (see discussion next paragraph).

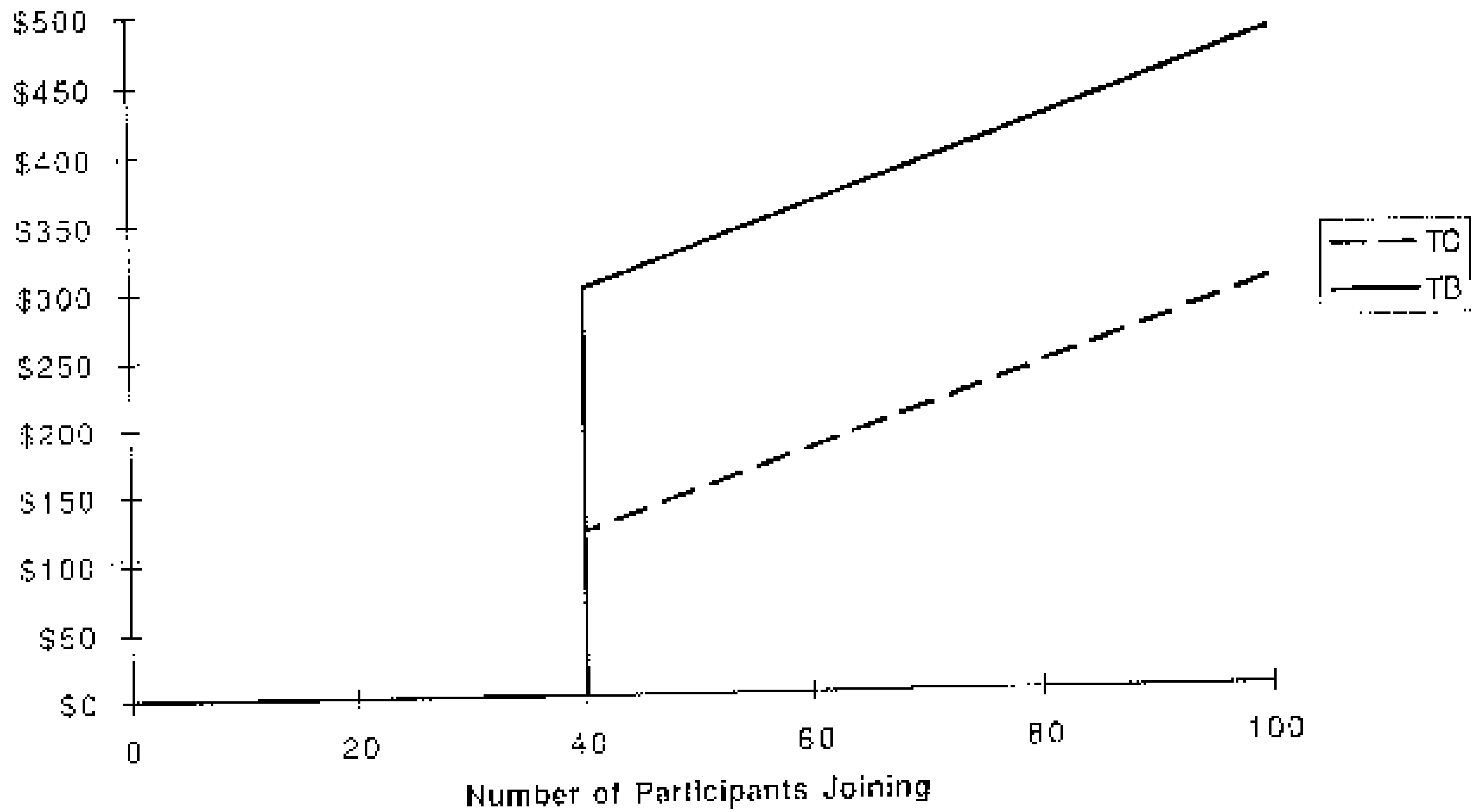
This section describes a classroom laboratory experiment specifically designed to evaluate the demand revelation properties of the NMPC mechanism. In addition to designing a laboratory mechanism paralleling the NMPC program, this experiment deviated from the body of previous public goods research in two important ways. First, in contrast to most public good experiments which have relied on "small groups" of less than 10 individuals, this experiment involved 100 participants. In part, this "large group" approach was adopted so as to more closely reflect the NMPC field conditions. The decision to use large groups was also based on experimental findings of Isaac, Walker and Williams (1994) that individuals in groups of 40 and 100 contributed significantly more to a VCM public good experiment than did subjects in small groups (n=4 and 10). Experimental results reported in Rondeau, Schulze, and Poe (1996) further suggest that a provision point mechanism using a proportional rebate conducted in a large group (n=45) setting approximates demand revelation in the aggregate while the same mechanism results in under-revelation in small groups (n=6). A second manner in which the analysis of the experiment contrasts with previous public goods research is that it models individual contribution decisions in a random utility framework.

The experiment was performed in an undergraduate economics principles class. The students had experience in market experiments but not in public goods experiments. An experiment "in decision-making" was introduced at the beginning of a regularly scheduled class, and printed instructions were distributed after students were seated. Students were instructed to copy the subject number written on their instructions onto a blank envelope which they were also provided. Students read their instructions (see sample in Appendix A), after which a brief oral summary was given. Questions were answered privately by monitors. Students were then allowed approximately ten minutes to make a decision which shall be described shortly. They then sealed their instructions and decision responses in their envelopes. Follow-up questions were distributed immediately afterward, and subject numbers were copied from the envelopes to follow-up questionnaires. All materials were collected after the follow-up forms were completed. The sealed envelopes ensured that students could not alter their decisions after answering the follow-up questions. Students were not allowed to communicate during the experiment.

The nature of the decision was as follows. Each participant was given a starting balance of \$5 and the opportunity to join a group investment program for a one-time fixed fee of \$3. Before a participant decided whether or not to join, the group investment program and payoff calculations were described. The group investment program would yield a return only if 40% or more of the participants joined. Each participant was informed that they would receive their pre-specified “return” if this provision point was met or exceeded regardless of whether or not they had joined. Each subject was randomly assigned to a return from the set {\$0.50, \$1.75, \$3.00, \$4.25, \$5.50}. Twenty subjects were assigned to each “return” and subjects were told their own return but were not made aware of the returns of other subjects. These returns were the induced values, designed to reflect the heterogeneous values NMPC customers hold for the *GreenChoice*™ program. If more than 40% joined, each participant also received a fixed "bonus payment" of 3¢ for each participant that joined in excess of the provision point. If fewer than 40% joined, the group investment program was canceled and all contributions were refunded. The bonus payment was public information.

The fixed participation fee was selected in conjunction with the induced values to insure that 1) the average payoff would equal or slightly exceed the participation fee and that 2) the total group benefits would equal or exceed twice the total group cost if the provision point were met or exceeded. Total costs (TC) and benefits (TB) are illustrated in Figure 1 for a group of 100 participants. This sample size was chosen to correspond with a large group setting, and to enable statistical analysis. The investment return values were chosen to be symmetric around the fixed fee and, based on pre-test results, to vary sufficiently to identify any relationship between induced value and participation for this sample size.

Figure 1: Total Costs and Benefits



44

The bonus mechanism was incorporated to reflect NMPC's offer of extended benefits financed by funds in excess of the provision point. The bonus amount of 3¢ was chosen so as to equate the aggregate group marginal benefits and marginal costs, as shown in Figure 1. The instructions were worded so as to avoid intrinsic value associated with program context; we sought to isolate the effectiveness of the mechanism alone in reducing free-riding behavior. Though this removed an important aspect of realism associated with NMPC's *GreenChoice*™ program, it allows for an unbiased evaluation of the program's financing mechanism. Finally, follow-up questions were posed to collect additional information on the participation decision (see Appendix B). The questions attempted to measure self interest and altruistic factors that might exogenously enter into participation decisions.

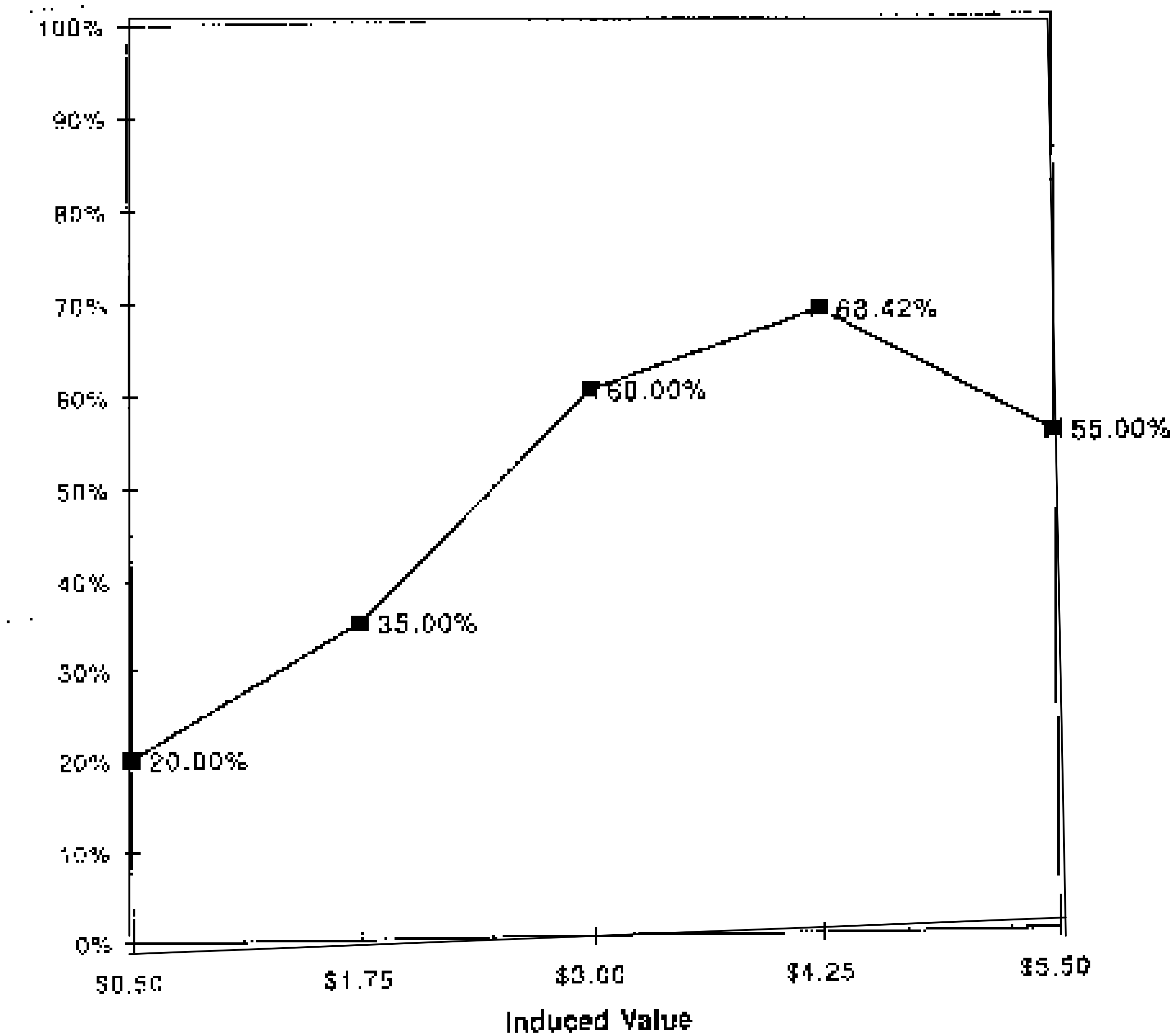
In summary, this experiment was designed to test the "naive" hypothesis that the provision point mechanism used by NMPC induces demand-revealing behavior under laboratory conditions. That is, we test if subjects with induced values above a posted price contribute and those with induced values below the posted price do not. If the mechanism is perfectly demand revealing, 50% of the 100 subjects should choose to participate in the program at a cost of $3, given the distribution of induced values: the 40% with induced values less than $3 should not sign up, the 40% with induced values exceeding $3 should sign up, and the 20% with the $3 induced value should be indifferent between joining and not joining. If, like the voluntary contribution mechanism, the provision point features fail to induce participation to levels approximating demand revelation, then we would expect that the results of the field experiment underestimate the "true" demand for the program.

4.2 Experimental Laboratory Results and Analysis

At the aggregate level, 47 percent of the subjects chose to join the program and pay the $3 fee, resulting in the funding of the public good. Clearly, this participation level closely approximates the 50 percent participation rate expected under our naive hypothesis. Thus, given this sample design, the mechanism appears to provide an approximately demand revealing outcome in the aggregate. In reaching this conclusion, it is interesting to note that in the week following the experiment described here, the same students participated in a standard computerized VCM public goods experiment developed by the Economic Science Laboratory at the University of Arizona. The experiment was conducted (using monetary incentives) as part of the students' regular weekly sections held in the Laboratory for Experimental Economics and Decision Research at Cornell. Contributions in the first round of this multiple round experiment were 41 percent of the maximum possible *payoff* (where the payoff corresponds to the induced value in the provision point experiment).[7] Thus, the subjects participating in these experiment appear typical, in that they exhibit substantial free-riding when in a single or initial period VCM environment (Davis and Holt, 1993).

As shown in Figure 2, participation is also generally responsive to increases in induced return. Contrary to our naive hypothesis, however, the response proportions do not exhibit a sharp step at $3. And thus, demand revelation associated with this mechanism is not perfect. Using a random utility framework first developed by McFadden (1976), it is possible to test the internal consistency of participation rates observed and the hypothesis that participation rates increase with induced value. In this framework, it is assumed that individuals know their own preferences with certainty, but that they may make errors in decision-making because of

Figure 2: Actual Joining Distribution
(By Induced Value)



imperfect information or errors in optimization. In addition, some aspects of the individuals' preferences are not observable by the analyst, and treated as random. These limitations introduce a stochastic error component into the modeling of decisions (Maddala, 1983).

Using such a model, we shall first specify the random utility equivalent of the naive null hypothesis, in which a customer will sign-up for the program at posted price \$C if the utility associated with having the program and paying \$C is greater than the utility associated with not having the program. If we assume that indirect utility is additively separable, the probability of a "yes" response to a particular posted price is then:

$$\text{(2)} \qquad \Pr\{\text{“Yes” response}\} = \Pr\{V - C + \epsilon > 0\}$$

where V is the value or willingness to pay of an individual for the green program and ϵ is an error term. Assuming that the error is logistically distributed, Equation (2) can be expressed as:

$$\text{(3)} \qquad \Pr\{\text{“Yes” response}\} = \frac{1}{1 + e^{-(\alpha + \beta(V - C))}}$$

where α and β are location and slope parameters to be estimated. The null hypothesis H_o^1: $\alpha = 0$ corresponds to the hypothesis that, at V = C, there is a 50 percent participation level. A positive value for α would shift the entire distribution to the left in a manner consistent with over-revelation relative to induced values, while under-revelation would correspond to $\alpha < 0$. The null hypothesis for the slope parameter H_o^2: $\beta = 0$ has only a one-sided alternative $\beta > 0$. That is, we are testing the hypothesis that participation does not increase with induced value.

Note from Equation (3) that for $\beta > 0$, the relationship between induced value and participation takes on an "S" shaped function through the introduction of logistically distributed random errors. Additionally, if $\alpha = 0$, when induced value equals cost (V = C), participation is

50%; as V-C becomes large, participation approaches 100%; and for small V relative to C, participation ultimately approaches 0%. The shape, or rather steepness, of the response function does vary with the magnitude of β. If $\beta = 0$, the probability of participation is a constant, but for large β, a step function is predicted. Figure 3 shows this relationship for a range of β values.

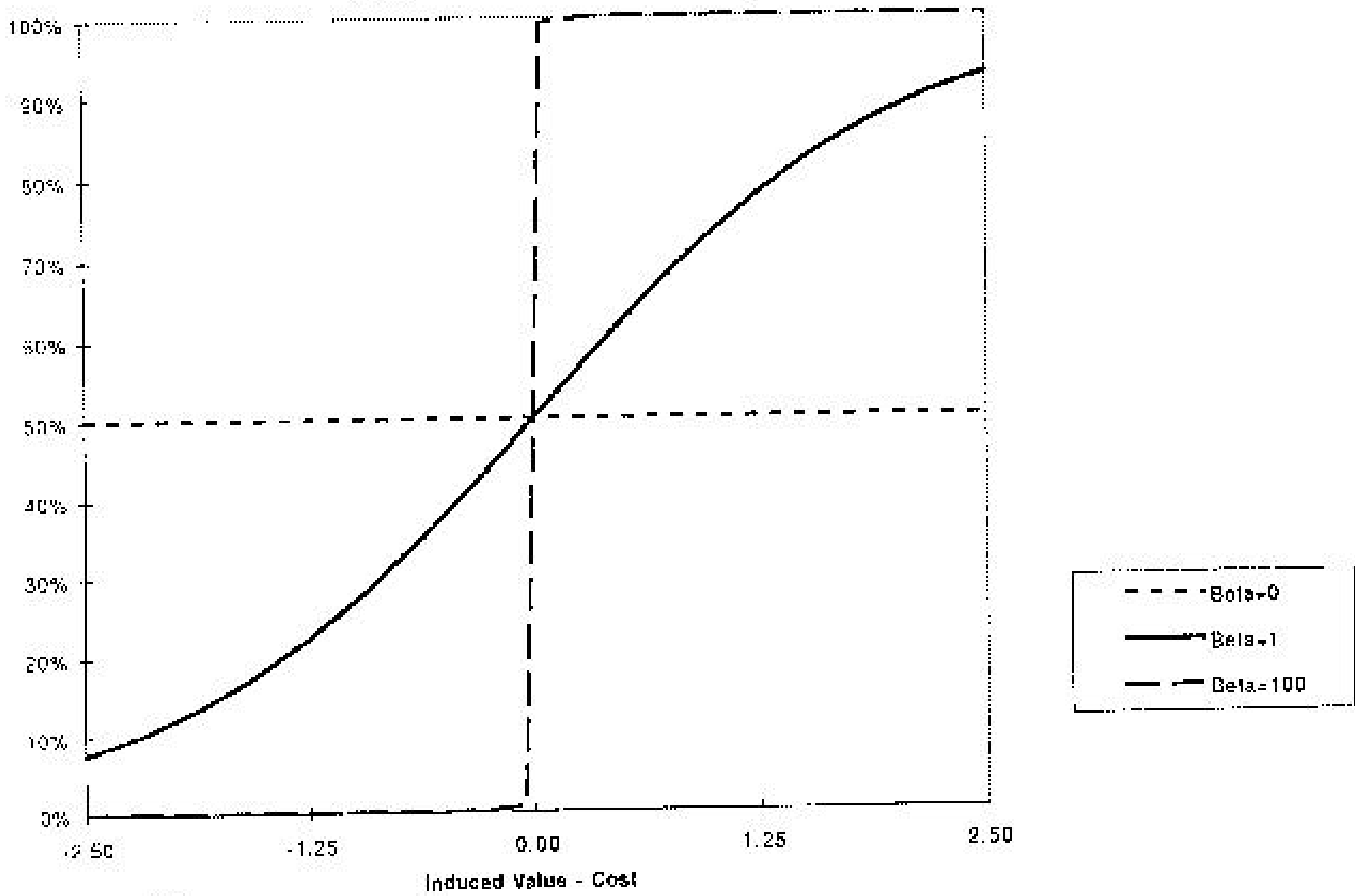
Estimates of α and β using maximum likelihood techniques are found in the "base" column of Table 2.[8] Consistent with our hypotheses, α is not significantly different from zero, indicating that the hypothesis of 50% participation at V-C = 0 cannot be rejected statistically. In addition, the estimated coefficient on V-C, β, is positive and significant. This latter result supports the hypothesis that participation is positively correlated with induced value. In all, these results are consistent with the hypothesis that this mechanism is demand revealing.

Table 2: Estimated Logit Models Using Induced Values

Variable (coefficient)	**Mean (s.d.) [Range]**	**Base**	**Long**
Constant (α_0)	1	-0.093 (0.211)	-2.26 (0.537)***
Group/Self (α_1)	0.61 (0.44) [0.14, 2.50]		3.688 (0.856)***
Induced Return (β)	0.01 (1.77) [-2.50, 2.50]	0.337 (0.123)***	0.301 (0.143)***
n		98	98
Likelihood Ratio χ_2		8.02***	38.19***
Percent Correctly Predicted		61	73

*,**,*** indicate significance levels of 10, 5, and 1 percent, respectively.

Figure 3: Random Utility Model for Various Betas
(By Induced Value minus Cost)



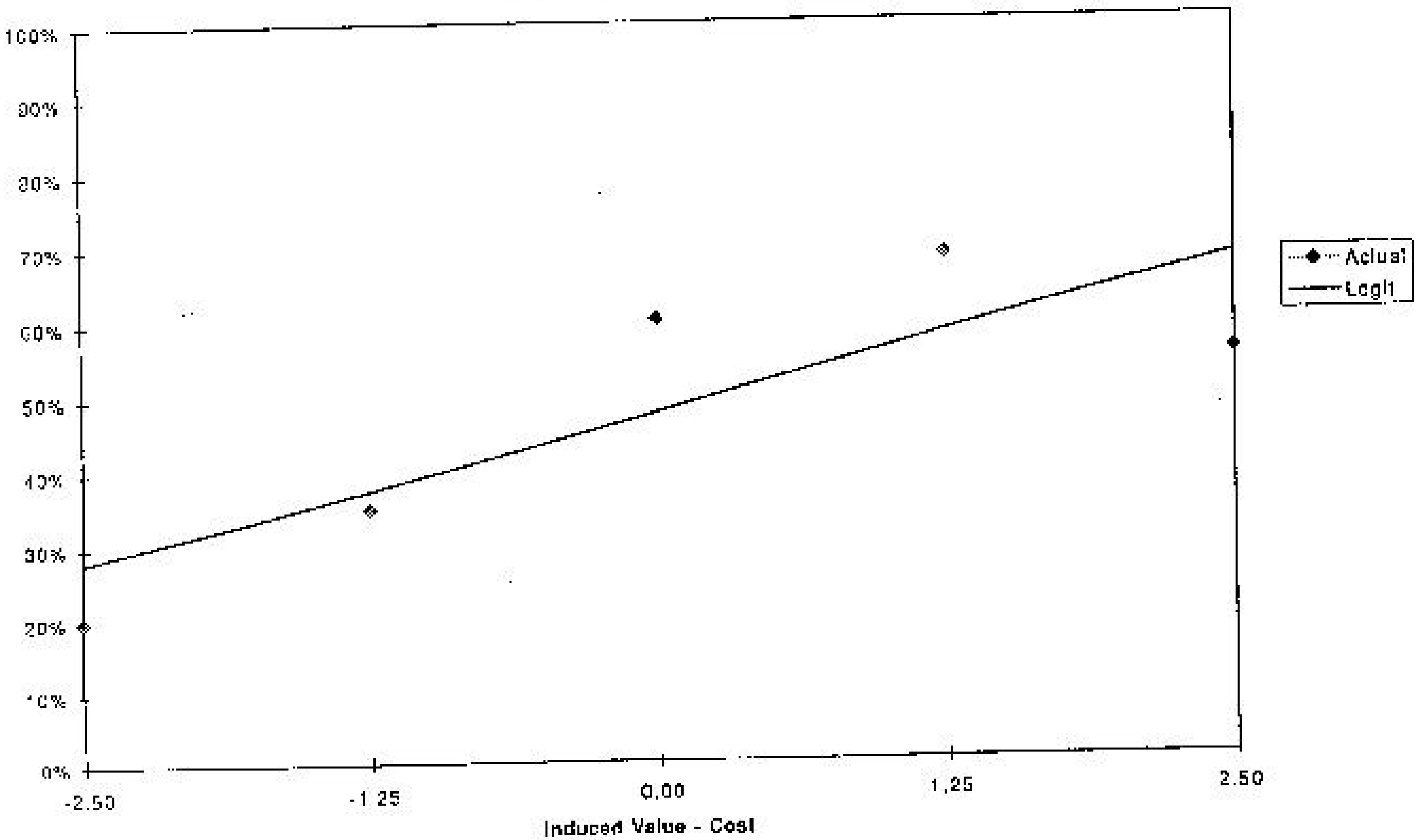
However, in spite of the highly significant estimation results reported in Table 2, closer examination of the data reveals that the model is not completely characterizing individual decisions. For example, as depicted in Figure 4, participation at lower values (e.g. V-C = -$2.50) exceeds the zero percent participation expected. There is also an obvious dip at the induced value of $5.50 (V-C = $2.50) The remainder of this section summarizes an exploratory investigation of why these deviations occur by focusing on altruistic and free-riding motivations. This extended analysis is intended, in part, to demonstrate the opportunities arising from a random utility modeling framework in future experimental economics research. The objective is to also provide an empirical base and motivation for future theoretical research.

An advantage of the random utility modeling is that it can incorporate other explanatory variables into the error based decision framework. In an effort to account for differential, exogenous motives, subjects were asked to indicate the importance they attached in making their decision to maximizing their own earnings, and to maximizing group earnings, both on seven-point scales (1 = Not Important, 7 = Extremely Important). Each of these questions are provided in Appendix B.

The self-reported interest in maximizing "group" and "self" earnings were combined in a "group/self" ratio so as to normalize relative responses at the individual level. In other words, a response pattern group=5, self=5 would be assigned a group/self ratio of 1, as would the response pattern group=2, self=2. In terms of Equation (3), this ratio (group/self) is included by expanding α from a constant to a vector and treating the ratio group/self as a separate element. As such, argument α in Equation (3) becomes $\alpha_{Grand} = \alpha_0 + \alpha_1$*(group/self). The expectation is that participation is positively related to group orientation, and thus α_1 should be positive with a corresponding null hypothesis H_0^3: α_1=0. To account for this ratio, the null hypothesis H_o^1: α=0, must be restated as H_0^4: $\alpha_{Grand} = (\alpha_0 + \alpha_1$*(group/self)) = 0. As before, a positive value for

Fig. 4

Figure 4: Actual vs. Logit Estimated Distribution (By Induced Value minus Cost)



α_{Grand} would shift the entire distribution to the left, indicating "over-revelation" associated with altruism. A negative α_{Grand} would shift the distribution to the right, providing evidence of free-riding.

The results from including this ratio in the estimation are provided in the "long" column of Table 2. Consistent with Andreoni's (1995) arguments concerning the role of altruism in public goods experiments, the estimated coefficient α_1 is positive and significant. Notably, the inclusion of this variable does not have a significant effect on the slope coefficient, but does greatly increase the explanatory power of the estimated model, as demonstrated by the jump in the percentage of responses correctly predicted and the likelihood ratio chi square values. Thus we argue that the addition of this variable makes a significant contribution to the explanatory power of the decision making model.

Setting the group/own ratio at its mean (0.61), α_{Grand} equals -0.01 (s.e. = 0.25) and is not significantly different from zero at any standard level of significance. As such the naive null hypothesis H_o^4: $\alpha_{Grand}=0$ still cannot be rejected for the average respondent in spite of the fact that the individual coefficients used in calculating α_{Grand} are each significantly different from zero. In other words, the altruistic behavior of subjects with induced values of \$0.50, \$1.75, and \$3.00, as captured by the positive and significant α_1 estimate, is being canceled out by the free-riding behavior of subjects with the higher induced values (recall Figure 2). It is interesting to note however that α_{Grand} is significantly different from zero in expected directions when the ratio group/self falls below 0.47 or exceeds 0.77. These results are consistent with previous research using split-sample designs to examine subject group effects in public good provision experiments, and provide additional evidence that participants bring different motives into experimental settings (Ledyard, 1995). From the perspective of this paper, these results in the "controlled environment" of the laboratory further heighten the importance of identifying respondent characteristics and preferences that may affect actual participation levels in field experiments.

5. Discussion and Conclusions

Green pricing programs have come under substantial criticism in the electric utility industry because of their cost and poor customer participation. Our field experiment shows that customers who are made fully aware of a green pricing program, and who face a provision point mechanism, participate at a relatively high rate (between 16 and 20 percent). The two completed programs in which provision points were utilized succeeded in funding local projects with relatively high levels of participation. Further, our laboratory examination of the NMPC mechanism found that it approached demand revelation both at the aggregate and individual level. These results suggest that the disappointing sign-up rates of most green pricing programs to date could well be due to free riding associated with mechanism design, as well as to the problem of limited customer awareness. It should be noted that it is difficult, time consuming, and expensive to raise customer awareness for new programs such as *GreenChoice*™. However, employing a provision point mechanism is a relatively costless way to increase participation. On a practical note, economists should recognize the large impediment that consumer awareness plays for the private provision of public goods. Our results suggest that the NMPC program may well have failed simply because the company was unable to expend sufficient resources to effectively market a statewide program. The successful provision point programs in Traverse City and Fort Collins funded local rather than statewide projects, so, given the high profile nature of wind energy projects, awareness was easily achieved. Finally, this research suggests that, where large groups are involved, provision point mechanisms may fulfill the objective of privately funding public goods.

APPENDIX A: Sample Subject Instructions for the Laboratory Experiment

Subject Number ____

<u>PRINT</u> your Name and Social Security Number so that we can pay you

Name ____________________________________

Social Security Number ____________________

INSTRUCTIONS

First, please write your subject number on the front of the envelope you have been given. You have been given the envelope to insure confidentiality.

This is an experiment in the economics of decision making. If you follow the instructions closely and make decisions carefully, you can earn money. Please do not communicate with any other students during the experiment. If you have any questions please do not hesitate to raise your hand so that someone can come over and answer your questions individually.

In this experiment all participants are given a starting balance of $5, which is yours to keep or use any way you like. At the end of these instructions, all of you will be asked if you want to join a group investment program for a one-time fee of $3. **The exact amount of money that you will earn in the experiment depends on your answer to this investment question, as well as on the answers of ALL the other participants in your group.** At the end of the experiment, your earnings will be calculated and you will be paid in cash.

Once you understand the group investment program and how your earnings will be calculated, your task is to decide whether or not you want to join the group investment program for a fixed fee of $3.

The group investment program works as follows. You are a member of a group of 100 people in this class. The program will only be funded and implemented if at least 40 of the 100 participants in your group join the investment program. If enough participants join the investment program so that the program is implemented, the return on the investment will be **SHARED BY ALL** participants in the experiment, **investors and non-investors alike**. Specifically, **regardless of whether or not you have joined the group investment program,** if enough people join, you will receive a return of $5.50. You will also receive a bonus payment of 3¢ for each participant that joins in excess of the minimum number of 40 necessary for the group program to be implemented. Furthermore, you keep your initial credit of $5 from which $3 will be deducted if you decide to join the investment program. <u>Note that other participants may have a different return but do **not** have a different bonus</u>.

If **not** enough participants join the investment program, the program will **not** be funded and will be canceled. In this case all the $3 fees collected will be refunded to those who joined. Thus, regardless of your decision to join the program or not, you would keep your $5 starting balance.

To Summarize:

- You must decide whether or not to join a group investment program for a cost of $3.
- If fewer than 40 participants out of 100 join, the program will be canceled and all $3 fees will be refunded.
- If 40 or more participants join, the program will be implemented and you will receive a return of $5.50 plus a bonus of 3¢ for each household that joins above 40.
- Recall, that you do not need to join to receive your payment from the investment program if 40 or more other participants join.
- But if you do join, you must pay the $3 fee.

This is the end of the instructions. If you have any questions please raise your hand.

THE QUESTION

Do you want to join the group investment program for a fixed fee of $3?
(Circle one only)

YES I wish to join

NO I do not wish to join

Please place this sheet in the envelope provided and seal it. When everyone has sealed their envelope, you will each be handed another sheet of questions. You must complete these additional questions in order to get paid.

APPENDIX B: Follow Up Questions for Laboratory Experiment

<u>TO BE PAID, YOU MUST COMPLETE THESE QUESTIONS</u>

Please enter your Subject Number from your envelope ___
<u>PRINT</u> your Name and Social Security Number as you did before
Name ____
Social Security Number ______________________

(1) Do you think that enough people joined to fund the group investment program?
(Circle one answer)

YES NO

(1a) More precisely, how many people do you think joined--excluding yourself?

____

(2) On a scale from 1 to 7, where 1 is not important and 7 is extremely important, how important were the following in your decision?

2a. I wanted to make as much money as I could for myself. (Circle one number)

1 2 3 4 5 6 7

Not Important Extremely important

2b. I wanted the group to make as much money as possible. (Circle one number)

1 2 3 4 5 6 7

Not Important Extremely important

1. We wish to thank the National Science Foundation, the Environmental Protection Agency and Niagara Mohawk Power Corporation as sponsors of this research. Specifically, we wish to acknowledgeTheresa Flaim, Janet Dougherty, Mike Kelleher, Pam Ingersoll, and Maria Ucchino at Niagara Mohawk Power Corporation. In addition, we benefited from valuable components from presentation discussants Martin Sefron and Andrew Platinga.

2. The authors are respectively: Research Assistant, Department of Agricultural, Resource and Managerial Economics (ARME), Cornell University; Visiting Professor, Department of Economics, University of British Columbia; Assistant Professor, ARME, Cornell University; Research Assistant, Department of Economics, Cornell University; Robinson Professor, ARME, Cornell University.

3. See Baugh et al. (1995) for a detailed discussion of Green Pricing Programs.

4. In a series of papers, Palfrey and Rosenthal (1984, 1988, and 1991) develop theoretical models of contributions to public goods when individuals face the binary choice of contributing either a posted price or nothing. Unfortunately, the complex environment under consideration in our experiment (a large group, heterogeneous valuations, and incomplete information about others' preferences) precludes a direct test of this theory. Note that Palfrey and Rosenthal analyze environments with homogeneous values, so demand revelation is not an issue.

4. In designing this program, NMPC asked William Schulze to suggest mechanisms to reduce free riding in green pricing programs (Schulze, 1994).

6. The survey instrument followed the Dillman Total Design Method for telephone surveys (Dillman, 1978) which is designed to achieve a high overall response rate by keeping text blocks short and clear and by engaging the respondent with frequent questions throughout the survey. The response rate was just under 70%. The survey was pretested by administering successive draft versions by phone until respondents clearly understood the instrument. Hagler Bailly Consulting, Inc. was contracted to administer the survey. Prior to telephone contact, potential respondents were sent a hand-signed cover letter on Cornell University stationery. The letter informed them that they had been selected as one of a small sample of customers to participate in the study of a new type of environmental program. It identified the study's sponsors as the National Science Foundation and the Environmental Protection Agency, together with NMPC, and enclosed a two dollar bill as a token of appreciation for participation. The two dollar bill has been found to be cost effective in increasing response rates.

7. Households were classified as "unable to contact" based on a minimum of eight attempts.

8. Respondents were also asked how they viewed the program in comparison with other causes they might support "*like the United Way, public television, or environmental groups,*" using a scale of one ("*much less favorably*") to 10 ("*much more favorably*") as a means of consolidating their preferences immediately prior to answering the participation question. Responses to this question are not included here, as they are a statistically significant function of the type of the

project as well as the mechanism attributes.

8. In the linear random utility model used in this analysis, income cancels out of the equation (Hanemann, 1984) and is thus not included here.

10. This contribution figure is based on 84 valid VCM observations from the same 100 students. The 16 invalid observations were due to computer malfunction, student absence, or untraceable student information data.

11. Only 98 observations are reported in Table 2, due to the fact that two respondents had missing values for various parts of the questionnaire.

REFERENCES

Andreoni, J., 1995. "Cooperation in Public-Goods Experiments: Kindness or Confusion?" American Economic Review 85(4): 891-904.

Bagnoli, M. and B. Lipman, 1989. "Provision of Public Goods: Fully Implementing the Core through Voluntary Contributions." Review of Economic Studies 56: 583-601.

Bagnoli, M. and M. McKee, 1991. "Voluntary Contributions Games: Efficient Private Provision of Public Goods." Economic Inquiry 29: 351-366.

Baugh, K., B. Brynes, C. Jones, and M. Rahimzadeh, 1995. "Green Pricing: Removing the Guesswork." Public Utilities Fortnightly, August: 26-28.

Brynes, B, M. Rahimzadeh, K. Baugh and C. Jones, 1995. "Caution Renewable Energy Fog Ahead! Shedding Light on the Marketability of Renewables." Paper presented at NARUC-DOE Conference on Renewable and Sustainable Energy Strategies in a Competitive Market, Madison, WI. May 1995.

Clements-Grote, 1997. Wind Power Pilot Program Information Packet. City of Fort Collins Light and Power Utility. Fort Collins, CO.

Cropper, M. L., and F. G. Sussman, 1990. "Valuing Future Risks to Life." Journal of Environmental Economics and Management 19: 160-174.

Davis, D. D. and C. A. Holt, 1993. Experimental Economics. Princeton: Princeton University Press. 325-333.

Dawes, R., J. Orbell, R. Simmons, and A. van de Kragt, 1986. "Organizing Groups for Collective Action." American Political Science Review 8: 1171-1185.

Dillman, D. A., 1978. Mail and Telephone Surveys - The Total Design Method, Wiley & Sons: New York.

Farhar, B. C. and A. H. Houston, 1996. "Willingness to Pay for Electricity from Renewable Energy." Paper presented at the 1996 ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA. August 25-31.

Groves, T., and J. Ledyard, 1977. "Optimal allocation of public goods: a solution to the 'free rider' problem." Econometrica 45, 783-809.

Holt, E. and Associates, 1996a. Green Pricing Newsletter. The Regulatory Assistance Project, Gardiner, ME. Number 3, April.

Holt, E. and Associates, 1996b. Green Pricing Newsletter. The Regulatory Assistance Project, Gardiner, ME. Number 4, October.

Holt, E. and Associates, 1997. Green Pricing Newsletter. The Regulatory Assistance Project, Gardiner, ME. Number 5, May.

Hanemann, W. M., 1984. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses." American Journal of Agricultural Economics 66: 332-341.

Isaac, R. M., D. Schmidtz, and J. M. Walker, 1989. "The Assurance Problem in a Laboratory Market." Public Choice 62: 217-236.

Isaac, R. M., J. M. Walker, and A. Williams, 1994. "Group size and voluntary provision of public goods: Experimental evidence using large groups." Journal of Public Economics 54:1-36.

Ledyard, J. O., 1995. "Public Goods: A survey of experimental research." In J. H. Kagel and A. E. Roth eds. Handbook of Experimental Economics. Princeton: Princeton University Press. 111-194.

Maddala, G. S., 1983. Limited Dependent and Qualitative Variables in Econometrics. New York: Cambridge University Press.

Marks, M.B., and R. Croson, 1996. "Alternative rebate rules in the provision of a threshold public good: an experimental investigation." Forthcoming in Journal of Public Economics.

McFadden, D., 1976. "Quantal Choice Analysis: A Survey." Annals of Economic and Social Measurement 5: 363-390.

Moskovitz, D. H., 1992. "Renewable Energy: Barriers and Opportunities: Walls and Bridges." Report for the World Resources Institute.

Moskovitz, D. H., 1993. "Green Pricing: Experience and Lessons Learned." The Regulatory Assistance Project, 177 Water Street, Gardiner, Maine 04345-2149.

Palfrey, T., and H. Rosenthal, 1984. "Participation and the provision of discrete public goods: A strategic analysis." Journal of Public Economics 24:171-93.

Palfrey, T., and H. Rosenthal, 1988. "Private incentives in social dilemmas: The effects of incomplete information and altruism." Journal of Public Economics 28:309-32.

Palfrey, T., and H. Rosenthal, 1991. “Testing game-theoretic models of free riding: New evidence on probability bias and learning.” Laboratory Research in Political Economy, T. Palfrey, editor, Ann Arbor: University of Michigan Press.

Poe, G. L., J. Clark, and W. D. Schulze, 1997. “Can Hypothetical Questions Predict Actual Participation in Public Programs: A Field Validity Test Using a Provision Point Mechanism,” Working Paper. Cornell University.

Rondeau, D., W. D. Schulze, and G. L. Poe, 1996. “Voluntary Revelation of the Demand for Public Goods Using a Provision Point Mechanism.” Submitted to the Journal of Public Economics. Cornell University.

Schulze, W., 1994. "Green Pricing: Solutions for the Potential Free-Rider Problem." Paper prepared for Niagara Mohawk Power Corporation.

Smith, V. L., 1979. “Experiments with a decentralized mechanism for public goods decisions.” American Economic Review 70, 584-599.

Suleiman, R. and A. Rapoport, 1992. "Provision of Step-Level Public Goods with Discontinuous Contribution." Journal of Behavioral Decision Making 5: 133-153.

Van Liere, K. D., and R. E. Dunlap, 1980. “The Social Bases of Environmental Concerns: A Review of Hypotheses, Explanations and Empirical Evidence.” Public Opinion Quarterly 44: 181-197.

Wood, L. L., W. H. Desvousges, A. E. Kenyon, M. V. Bala, F. R. Johnson, R. Iachan, and E. E. Fries, 1994. “Evaluating the Market for ‘Green’ Products: WTP Results and Market Penetration Forecasts,” Working Paper #4. Center for Economics Research, Research Triangle Institute.

CHAPTER 4
Can Hypothetical Questions Predict Actual Participation in Public Programs? A Contingent Valuation Validity Test Using a Provision Point Mechanism

1. Introduction

A critical issue in environmental economics and public policy is the ability of contingent valuation (CV) to measure "actual" willingness to pay (WTP) for environmental commodities (Arrow *et al.*, 1993). Early validity field research compared hypothetical CV responses with values obtained from auctions and other actual market transactions for private (e.g. strawberries, Dickie *et al.*, 1987) and quasi-public goods (e.g. hunting permits, Bishop and Heberlein, 1979), concluding that "the overwhelming weight from simulated market experiments favors the use of contingent valuation for estimating willingness to pay" (Bishop and Heberlein, 1990). More recent efforts have sought to extend the CV/actual market comparisons to less familiar public goods with large nonuse components: Duffield and Patterson (1992) conducted such comparisons for leasing water rights for threatened trout streams, Seip and Strand (1992) evaluated hypothetical and actual sign-ups for an environmental organization, Brown *et al.* (1996) compared hypothetical and real donations for the removal of roads on the north rim of the Grand Canyon, and Navrud and Veisten (1997) compared hypothetical and actual payments for old growth forest parcels. Together, these studies have suggested that there are considerable differences between hypothetical and actual contributions, which have largely been attributed to biases associated with the hypothetical nature of CV. For example, Brown *et al.* (1996, p. 164) argue that "Hypothetical questions, especially about donations to generally desirable environmental goods seem to engender overestimates of actual WTP." Such conclusions have lent some support to efforts to discredit CV as a public policy decision tool. They

***Authors: Gregory L. Poe, Jeremy E. Clark, and William D. Schulze**

have also led to efforts to "calibrate" hypothetical CV responses to better approximate measures of actual WTP (Champ *et al.*, 1995, 1997).

Using these past field comparisons as a justification for not using CV in public policy applications, or as a metric for calibrating CV values, is inappropriate, as each of these comparisons relies on a voluntary market contributions mechanism (VCM) as a criterion for conducting the validity test. Theoretical developments following Samuelson (1954) and decades of experimental economics research indicate that these mechanisms are neither incentive compatible in theory nor demand revealing in practice (see Ledyard, 1995, for a comprehensive review of the literature): i.e. free riding is the expected norm. In VCM experiments involving real money, individuals typically contribute 40 to 60 percent of the Pareto optimum level. This failure of the VCM to elicit honest revelation of demand for public goods in the laboratory makes it an inappropriate market criterion to assess the validity of CV in field tests. Importantly, it casts serious doubts on the claim made by other researchers that CV suffers from hypothetical bias. Indeed, it is possible that the difference between hypothetical WTP and actual contributions found previously could largely be attributed to free riding rather than an expression of upward bias in hypothetical answers.[1]

In order to conduct more accurate field validity tests of the contingent valuation method, a demand revealing one-shot mechanism is needed to collect actual payments that more closely reflect

---

[1] To their credit, each of the aforementioned authors are apparently aware of the possible biases associated with using a VCM as a reference criterion for willingness to pay. For example, Seip and Strand note, "We may have significant free rider problems in voluntary payment" (p. 103). Brown *et al.* similarly note that " A voluntary payment towards a public good allows for free-riding...to the extent that free-riding occurs, it depresses actual payments" (p. 154). Yet, the explicit or implicit justification (e.g., Champ *et al.*, 1997) in these studies that such a contribution provides a lower bound of value belies the critical point that such "lower bound" measures do not provide the Hicksian measures appropriate for welfare economic analyses.

true WTP. Here, demand revelation is defined in the purely empirical sense that individuals provide their true values through payments to a public fund. Importantly, the mechanism needs to be demand revealing in a single request for contributions so that it mirrors the context employed in actual CV. Building on recent experimental economics research, this paper presents the first field study to use an approximately demand revealing reference mechanism to test hypothetical bias and possible calibration alternatives in contingent valuation.

This research is particularly timely because it focuses on green pricing, an approach increasingly being considered by utilities to fund the development of renewable energy resources.

> *"Green pricing is a generic term for the offer of electricity generated from clean, environmentally preferred sources such as solar, wind, geothermal and some types of biomass and hydro energy sources. Consumers who choose to purchase this product pay a small premium for the green electricity. This idea has been getting significant attention since its conception in 1992 (Moskovitz, 1992). Seven utilities now have some form of green marketing program in operation, and some twenty others have been considering whether to offer green pricing, including conducting market research into consumer preferences."* (Holt, 1997, p. 1)

Utility interest in this product has been motivated, in part, by national environmental opinion polls suggesting that majorities are willing to pay $6 to $25 more per month for green power (Farhar and Houston, 1996; Holt, 1997)[2]. Despite substantial evidence supporting public interest in green pricing programs, actual programs have been characterized by low participation rates, usually less than 2

---

[2] Specific market research conducted for individual utilities generally supports these broad-based attitudinal findings: a Sacramento Municipal Utility District (SMUD) study indicated that 43 percent of residential customers would pay 5 percent or more for their energy for SMUD to invest in renewable energy products; a Massachusetts Electric Company survey estimated that 48 percent of the residential customers would probably or definitely sign up at a 5 percent premium; Niagara Mohawk Power Corporation research suggested that 60 percent of residential customers would sign up at a fixed rate of $6 per month; and survey research for an unnamed western electric utility suggested that 82 percent of residential customers would be willing to pay rate premiums for renewable energy generation programs (Holt, 1997; Brynes *et al.* 1995)

percent (Farhar and Houston, 1996).

Three explanations for this discrepancy seem likely. First, most utility customers may have been unaware of such programs, in spite of attempts by electric utilities to inform them using bill inserts, mailed brochures, and advertising (Farhar and Houston, 1996; Holt, 1997). Since market research must necessarily inform customers of a green pricing program, it inherently creates perfect awareness. Thus, forecasts derived from market research depend critically on assumptions about the effectiveness of marketing (Wood *et al.*, 1994). Second, free-riding may lower customer participation because participation has usually been structured as a charitable voluntary contribution.[3] Finally, the hypothetical questions used in market research studies may lead to upward bias. In this paper we control for the first two issues by using a phone survey to create 100 percent awareness and utilizing an approximately demand revealing mechanism for the actual collection of funds in a GreenChoice™ program offered by Niagara Mohawk Power Corporation (NMPC) developed with input from one of the authors (Schulze, 1994). Thus we are able to isolate a direct test of

---

[3] Interestingly, the other green pricing programs to use provision point mechanisms have met their (often fairly low) funding objectives in contrast to early "unsuccesssful" programs associated with voluntary contribution mechanisms. Traverse City Light and Power and Fort Collins Light and Power attempted and completed windmill projects using a first-come-first-serve provision point mechanism. Participation in the Traverse City project was curtailed after the program's provision point was successfully reached with 263 customers, at an estimated 3.4 percent of targeted customers. At the time of this paper, Fort Collins had funded one of three windmills with no participation rates reported. Unfortunately, though the Niagara Mohawk Power Corporations GreenChoice™ program was formally approved by the New York Public Service Commission, it was ultimately suspended before completion because NMPC developed serious financial difficulties and was unable to promote customer awareness of the program. Before suspension, the program was briefly mentioned in a bill insert and described in a brochure sent to about 38,000 of Niagara Mohawk Power Corporation's 1.2 million customers. Most of the planned marketing campaign, including a substantial advertising budget and tree plantings at public schools throughout the service territory, was canceled, and thus awareness was extremely low. Before the program was terminated, however, we were able to conduct the field experiment described in this paper using a phone survey of NMPC customers.

hypothetical bias associated with different contingent valuation elicitation methods for this green pricing program.

The remainder of this paper is organized as follows. The next section provides a review of demand revealing public goods mechanisms and development of the mechanism used in this research. The third and fourth sections describe the results of the validity test and the assessment of calibration tools for open-ended and dichotomous choice contingent valuation. Conclusions and implications are provided in the final section.

2. THE SEARCH FOR DEMAND REVEALING PUBLIC GOODS MECHANISMS

An important finding from decades of experimental economics research is that no public goods elicitation mechanism, even if it is theoretically incentive compatible, is perfectly demand revealing in practice (Smith, 1979, 1980; Harstad and Marrese, 1982). Some public goods mechanisms that increase incentives for honest revelation have been developed over the years (e.g. Groves-Ledyard, Tullock, Smith Auction, Clark), and have been shown to approach demand revelation in laboratory experiments in which induced values can be compared to actual contributions. Extending these mechanisms to CV field research is, however, problematic. Mechanisms such as the Groves-Ledyard (Groves and Ledyard, 1977) often involve extremely complex incentive structures -- greatly limiting their applicability and usefulness in situations outside of confined laboratory settings. Other mechanisms, such as the Smith Auction (Smith 1979, 1980; Coursey and Smith, 1984; Harstad and Marrese, 1982), require unanimity, which necessitates an interactive small-group situation. Moreover, these techniques require multiple rounds before they approximate group demand revelation, and, thus are not readily applicable in one-shot situations. Thus, it is understandable that prior CV public goods field validity studies (Duffield and Patterson, 1992; Seip and Strand, 1992; Brown *et al.*, 1996; Champ *et al.*, 1995, 1997; and Navrud and Veisten, 1997) have relied on voluntary contribution mechanisms to collect actual payments. However, as noted above, the voluntary contribution mechanism (VCM) is not theoretically incentive compatible and not empirically demand revealing.

In recent research, three important modifications to the VCM have been shown to reduce free-riding in public goods experiments: 1) A provision point (PP) is a minimum level of aggregate contributions below which the public good is not provided. Isaac, Schmidtz and Walker (1989), Suleiman and Rapoport (1992), and Dawes, *et al.* (1986) report that adding a PP significantly increases contributions in a variety of experimental environments. 2) A money-back guarantee (MBG) can be added to a provision point mechanism (PPM) so that individual contributions are refunded if the PP is not reached by the group. Isaac, Schmidtz and Walker (1989) report that contribution levels were substantially and significantly higher in treatments with the MBG compared to baseline PP experiments without a refund. 3) A rebate rule for disposing of contributions in excess of the provision point is a second form of assurance against the potential loss of contributions. Rebate rules can take various forms (see Marks and Crosson, 1998, for a study of alternative rebate rules). For example, in the proportional rebate (PR) rule, all excess contributions are returned to contributors, with each individual receiving a share of the excess proportional to the weight of her contribution to the group fund. In an extended benefits (EB) rule, money collected in excess of the provision point would be used to extend benefits, or increase the production of the public good. Extending benefits beyond the provision point does not modify individual incentives in theory, but simply creates a VCM environment beyond the threshold. In evaluating alternative rebate rules for provision point mechanisms experimentally, Marks and Crosson (1998) found that offering extended benefits had the greatest positive effect upon group contributions.

In all, the PPM with MBG and PR or EB rebate rules have been shown to be successful at increasing contributions to public goods in experimental settings. There are also anecdotal reports of provision points being used to successfully resolve actual free-riding problems (e.g. Bagnoli and McKee, 1991). Yet these mechanisms typically fail to produce demand revelation. Rather, as shown theoretically by Bagnoli and Lipman (1989) and suggested in empirical applications by Bagnoli and McKee (1991), these mechanisms produce an infinite number of Nash equilibria where the sum of contributions exactly covers costs (PP). In these circumstances efficiency is achieved but demand is under revealed (with the exception of the case where cost equals willingness to pay). It should be noted, however, that all existing PPM experiments have been conducted in experimental settings that greatly depart from field conditions: in particular they involve small groups with multiple rounds. In an effort to create a closer parallel with single shot, large group situations relevant for CV field research, we conducted a series of provision point experiments which are briefly reviewed here and reported in detail in Rondeau *et al.* (forthcoming) and Rose *et al.* (1996).

The Rondeau *et al.* (forthcoming) experiments tested the demand revelation characteristics of a one-shot PPM with MBG and PR in experimental environments selected to mimic continuous contribution CV field conditions. In this series of experiments, students with heterogenous induced values were given the opportunity to participate in a group investment fund by contributing all or a portion of their endowment. If the sum of the group contributions exceeded a predetermined threshold (the provision point), then individuals were paid their induced value along with a proportional rebate of excess contributions. If, instead, the provision point was not met, then the contributions were refunded. As a control, this mechanism was conducted in “small group” (n=6) experiments, resulting in an average ratio of revealed to induced demand of 67 percent. These results are far short of demand revelation but consistent with aggregate demand results found in previous PPM experiments (64 percent, Marks and Crosson, 1996; 61 percent, Cadsby and Maynes, 1996; 79 percent, Bagnoli and McKee, 1991). Using this experiment as a base, and noting Isaac,Walker, and Williams’ (1994) finding that individuals in groups of 40 and 100 participants contributed significantly more to a VCM public good than did subjects in small groups, a series of “large group” experiments with 45 to 50 participants was conducted. In contrast to the small group experiments, the results from this set of experiments suggest that the PPM/MBG/PR mechanism is approximately demand revealing in the aggregate when used in large groups of subjects who have heterogenous valuations for a public good. These results are shown to be robust to subject type and the level of information provided about the group’s size and provision point: participants revealed approximately 95 to 112 percent of induced demand across a range of treatments. In contrast to the small group results, each of these values is not statistically different from 100 percent demand revelation.

Based on the promising nature of these results, a second research effort was undertaken to examine a mechanism that more closely corresponded to the actual NMPC collection mechanism (see Rose *et al.*, 1996). The specific PPM design we considered corresponds to that used by NMPC to accelerate the development of renewable energy sources of electricity. The mechanism adopted by NMPC employed three of the features discussed previously. First, it contained a provision point of $864,000 to be raised through customer contributions. This minimum level of funding would provide for the construction of a renewable energy facility to serve 1,200 homes, and for the planting of 50,000 trees in the NMPC service area. Second, NMPC's funding mechanism offered a money back guarantee to customers which assured them that, if contributions failed to reach the threshold, all money collected would be refunded. Third, the mechanism offered the possibility of extended benefits. Money collected in excess of the provision point would be used to increase the number of homes served with renewable energy or to plant more trees.

One theoretically undesirable feature of NMPC's PPM/MBG/EB mechanism was that, to legally qualify as a rate offering, the program could only be offered at a single posted price. Thus, customers could choose only to contribute a fixed amount of $6.00 per month or not participate at all. A posted price is undesirable because it does not allow households to self-select a monthly fee that better represents their preferences for the program. Note that, despite the posted price, the mechanism does not reduce to a referendum, because the only individuals to pay are those who choose to participate.

A laboratory experiment was designed to test the NMPC funding mechanism in a large group environment where program values could be induced. The experiment was performed in an undergraduate business-economics principles class using 100 students. This subject pool was specially chosen to give the mechanism a rigorous test because business and economics students contribute less in VCM experiments than other groups (Cadsby and Maynes, 1996). Each participant was given an opportunity to join a group investment program for a one-time fixed fee. The group investment program only yielded returns if there were at least 40 participants in the investment fund (the PP). Five groups of 20 subjects were assigned to one of five induced values ranging from 17 to 183 percent of the fixed fee. If enough subjects joined, each of the 100 participants received a "bonus payment" (the EB), equivalent to 1% of the fixed fee, for each participant that joined in excess of the provision point. If fewer than the provision point joined, the group investment program was canceled and all contributions were refunded (the MBG).

The experiment was designed such that, if subjects behave as if the NMPC mechanism is demand revealing, we would expect that 50 percent of the subjects would choose to participate in the program. We would also expect that, under a random utility model that accounts for individual errors, the participation rate should increase with induced value. At the aggregate level, 47 percent of the subjects did chose to participate, resulting in the Pareto optimal funding of the public good. This participation level closely approximates the 50 percent participation rate predicted for demand revelation. Thus, in aggregate terms, this mechanism appears to provide an approximately demand revealing outcome for this sample design. Furthermore, using a random utility based logistic model, the response rate significantly increases with induced values and the predicted response is not significantly different from 50 percent participation at the point where induced value equals cost. In conjunction with the demand revelation characteristics of provision point mechanisms reported in Rondeau *et al.* (*forthcoming*) this research suggests that, in a single-shot large group setting analogous to that used in contingent valuation validity testing, provision point mechanisms approximately reveal demand in the aggregate.

3. THE NIAGARA MOHAWK POWER CORPORATION FIELD VALIDITY TEST

3.1. Experimental Design

Given that the NMPC mechanism appears to be approximately demand revealing in the aggregate in laboratory experiments, we maintain that it provides a substantially better baseline than the VCM for a field validity test of contingent valuation. On this basis, we applied this mechanism in a telephone survey comparison of hypothetical and real commitments to NMPC's GreenChoice™ program. As described previously, the program used a PP/MBG/EB mechanism, with a single posted price of $6 per month. Contingent valuation responses were collected using two telephone formats. The first was a dichotomous choice version directly paralleling the actual solicitation. The second was an open-ended version asking respondents the most they would be willing to pay for the program. These two survey formats offer extremes on the continuum of continuous to discrete choice contingent valuation. Past experimental economics and contingent valuation research have demonstrated that substantial procedural variance exists between these formats (see summaries in Brown *et al.*, 1996, and Schulze *et al.*, 1996). A critical question from a policy standpoint is which format most closely approximates actual preferences. We will be examining this question in the case of a public good offered at a single price.

All survey instruments followed the Dillman Total Design Method for telephone surveys (Dillman, 1978). The method generally achieves a high overall response rate, emphasizing short, clear text blocks and engaging respondents with evenly spaced questions throughout the survey. The program description was designed to correspond to the actual NMPC solicitation materials distributed to the public, despite the fact that these materials provided substantially less information than state of the art in contingent valuation research. In order to control for awareness, phone, rather than mail, surveys were employed.

Successive pretests of the survey were administered by phone to ensure that respondents clearly understood the instrument. The final phone survey was administered by Hagler Bailly Consulting, Inc., using a random sample of households in the Buffalo, NY area. Households in the sample were first sent a hand-signed cover letter on Cornell University letterhead announcing the survey. The letter informed them that they had been selected as one of a small sample of customers to participate in the study of a new type of environmental program. The study's sponsors were identified as the National Science Foundation and the Environmental Protection Agency, together with NMPC. A two dollar bill was enclosed as a token of appreciation for participation.

The phone survey itself ran as follows. Both actual and hypothetical versions began by reaching the person in the household who usually paid the NMPC electric bill. Speaking to that person, the interviewer described the survey's purpose and sponsors. The individual was then asked to rate NMPC's service. Next, customer awareness of the program was obtained, and the goals of the program were described. Respondents were then asked about their interest in these objectives:

> *How interested are you in the goal of replacing fossil energy with renewable energy sources? On a scale from 1 to 10, where 1 is not at all interested and 10 is very interested, how interested are you?*

and later:

> *How interested are you in the goal of planting trees on public lands in upstate New York? As before on a scale from 1 to 10, where 1 is not at all interested and 10 is very interested, how interested are you?*

Depending on the version, the funding plan was then described as follows:

> *The GreenChoice program would be funded voluntarily. Customers who decide to join the program would pay an additional fixed fee of $6 per month on their NMPC bill. This fee would not be tax deductible. Customers could sign up or cancel at any time. While customers sign up, NMPC would ask for bids on renewable energy projects. Enough customers would have to become GreenChoice partners to pay for the program. For example if 12,000 customers joined the first year, they would invest $864,000, which would allow Niagara Mohawk to plant 50,000 trees and fund a landfill gas project. The gas project could replace all fossil fuel electricity in 1,200 homes. However, if after one year, participation were insufficient to fund GreenChoice activities, Niagara Mohawk would cancel the program and refund all the money that was collected.*

For the open-ended format, the underlined section was removed. The exact dollar amount of the provision point was hedged somewhat by NMPC so that the renewable energy project could be sent for competitive bid while the program was underway.

The survey then asked respondents whether the program's funding mechanism made them more or less interested in the program. After this, respondents in the actual version were faced with the participation question:

> *So far I've described the GreenChoice program, as well as the $6 per month cost it would add to your household's electrical bill, if you were to join. You may need a moment to consider the next couple of questions. Given your household's income and expenses, I'd like you to think about whether or not you would be interested in the GreenChoice program. <u>If you decide to sign up, we will send your name to Niagara Mohawk, and get you enrolled in the program. All your other answers to this survey will remain confidential. Does your household want to sign up for the program at a cost of $6.00 per month?</u>*

In the hypothetical dichotomous choice version, the underlined portions were replaced by: "Would your household sign up for the program if it cost you $6 per month?"

The hypothetical open-ended decision question was also worded in typical fashion:

> *So far I've described the GreenChoice program. You may need a moment to consider the next couple of questions. Given your household's income and expenses, I'd like you to think about whether or not you would be interested in the GreenChoice program. What is the highest amount, if anything, that your household would pay each month and still sign up for the program?*

All surveys ended with debriefing and socio-economic questions useful for modeling demand.

As it turned out, contributions in the actual version were never collected, because the GreenChoice™ program itself was canceled. NMPC developed severe financial difficulties, and, having failed to pay dividends to stockholders, was unable to advertise the GreenChoice™ program. Consistent with the money back guarantee, those who elected to participate as a result of our phone survey were sent a cancellation notice, and the funds contributed by all households who signed up were returned. It is, of course, possible that the customers that we signed up might have reneged by leaving the program during the 12 month payment period. However, there is early evidence that this is not a large issue. For example, 95 percent of the residents who signed up for the Traverse City Wind Power project continued to pay their committed level more than one year after the program started. With actual and hypothetical measures of participation identified, we turn next to the results of the surveys.

3.2 Results and Analysis

A random sample of 1250 households in the Buffalo, NY area, based on zip code delineation, was purchased from a marketing research firm. An adjusted sample of 985 households remained after removing bad addresses, unlisted numbers, non-NMPC customers, and three respondents who had previously heard of the GreenChoice™ program. Among these 985 households, 206 were in the actual mechanism sample, 393 were in the hypothetical open-ended sample, and 386 were in the hypothetical dichotomous choice sample.[4] Of the total adjusted sample, 177 refused to participate, yielding an overall response rate of 72.5 percent. None of the subsample response rates fell below 70 percent.

---

[4] A modified, shortened "Cheap Talk" warning (Cummings and Taylor, 1996) was used on a sub sample of each of the hypothetical surveys in an effort to "push down" any hypothetical bias. This was found not to make a difference, which corresponds with the findings reported by Cummings and Taylor. In their research a lengthy version of "Cheap Talk" did, however, lower stated values. Our "Cheap Talk" versions were pooled with responses from surveys without this warning in the analyses in the text. Specifically, the following wording was added to the paragraph preceding the hypothetical valuation question *"I have one caution though. For programs like this it's often the case that more people say they would sign up than actually do sign-up. Utilities in other parts of the country have found that eight times as many people say yes to similar programs as actually take part in them. With this in mind..."*

Of the actual mechanism sample of 206, 179 were reached by phone. Of these, 37 refused or did not complete the survey. Of the remaining 142 respondents, 29 signed up for the program, resulting in a participation rate of 20.4 percent. Participation would fall to 16.5 percent if we assume that the 37 people contacted who did not complete the survey would have declined the program. Note also, only three people from our entire sample recalled having heard about the program, reflecting NMPC's decision not to market the program. As such, these data indicate strong potential support for the GreenChoice™ program amongst NMPC customers, and suggest that the program could have been funded if marketing had been successful in increasing awareness.

The participation rate of 16-20 percent is also substantially higher than that observed in the majority of other green pricing programs reported in the literature (Baugh *et al.*, 1995; Brynes *et al.*, 1995; Holt, 1997; Farhar and Houston, 1996). There are, however, substantial differences between this and most previous programs. First, program awareness was controlled here at 100%. In previous programs, participation rates have typically been defined over the broader base of total customers or customers targeted with direct mailings. Yet, as our findings suggest, customer inserts and direct mailings do not guarantee even minimal awareness among customers. Secondly, as noted, previous participation programs have mostly relied upon voluntary contributions, rather than the provision point mechanism used here.

Thus, the 20.5 percent sign-up rate provides a benchmark for testing the hypothetical bias associated with open-ended and dichotomous choice CV questions among survey respondents. We do so using two methods of analysis. First we compare participation rates across actual and hypothetical versions using simple Contingency Table analyses (Conover, 1980). Second, we model the participation decision and, controlling for socio-economic and other factors, test the hypothesis that participation rates differ between actual and hypothetical treatments. To conduct the analysis, open-ended responses are converted to participation rates based on whether the values given exceed the $6 threshold.

As shown in Table 1, the estimated participation rate from the open-ended responses is 23.9 percent, or 17 percent higher than the actual participation rates. The 30.6 participation rate from the dichotomous choice responses is 50 percent higher than the actual participation rates. These results contrast with the NOAA panel recommendation that dichotomous choice values offer conservative, and thus preferable, estimates of value (Arrow *et al.*,1993). At the same time, they are consistent with previous comparisons in CV and laboratory experiments (see Brown *et al.*, 1996 and Schulze *et al.*, 1996 for recent reviews). A Chi-Squared contingency table analysis rejects the hypothesis that actual and hypothetical open-ended participation levels differ at any standard level of significance ($\chi^2 = 0.66 < \chi^2_{1,\,0.10} = 2.71$). In contrast, the dichotomous choice sign-up rate of 30.6 percent is higher than the actual value of 20.4 percent at the 10 percent level of significance ($\chi^2 = 4.83$).

Following established dichotomous choice valuation techniques, we next assume a logistic distribution function to model an individual's participation decision as a function of covariates elicited in the questionnaire. Using actual responses as a base, we include binary variables to test whether each of the hypothetical patterns is significantly different from the actual response pattern.

Three categories of covariates are included when modeling participation. The first concerns respondents' support for the particular objectives of the program: replacing fossil fuels and planting trees in upstate New York. Interest in each goal was measured using a scale of one ("not at all interested") to 10 ("very interested"). Both scale responses are expected to be positively correlated with participation.[5]

The second category of covariates includes demographics, such as gender (male=1), age (in years), and education (college graduate or higher =1). Also included here are recent financial support of environmental groups (Yes=1), and impression of the overall service received from NMPC on a 1 ("very poor") to 10 ("very good") scale. These types of variables are widely used as explanatory covariates in the literature modeling environmental valuation.[6] From this literature we expected age to be negatively correlated with participation, and education, impression of NMPC service, and participation in environmental groups to be positively correlated with participation. No sign expectation was formed for gender.

The final category of covariates concerns respondents' views of the program's funding mechanism. These variables are unconventional, in the sense that they do not proxy for the value of the program itself. When told of the provision point and money back guarantee, respondents were asked the following two questions:

> *Does the fact that a minimum level of customer participation is required for GreenChoice to operate make the program of less interest to you, more interest, or does it not affect your interest?*

---

[5] Respondents were also asked immediately prior to the participation question how they viewed the program in comparison with other causes that they might support, like the United Way, public television, or environmental groups, again using a scale of 1 (much less favorably) to 10 (much more favorably). This question was included to allow participants to consolidate their preferences and perceptions, as well as to remind them of possible substitutes for this program. Responses are not included in the econometric analysis because individual responses were endogenous functions of expressed interest in program characteristics and mechanism design.

[6] The estimation procedure was motivated by a linear random utility difference model. Thus income is not included in the estimation (Hanemann, 1984). Similarly, in contrast to standard dichotomous choice CV models, price is not included as an explanatory variable of participation, because it is constant at $6 across all participants.

and,

> *Does the fact that Niagara Mohawk would refund all the money it collects--if support is insufficient--make GreenChoice of less interest to you, more interest, or does it not affect your interest in the program?*

The provision point itself did not arouse greater interest in the program. Over 55 percent responded that its inclusion did not affect their interest. Only 17 percent indicated that it increased their interest. In contrast, the money back guarantee increased interest in the program for 47 percent of respondents. Only 9 percent said that it reduced their interest. Both questions were recoded as binary variables for estimation, assigned ‘1' for "more interest," and ‘0' otherwise. We expected their coefficients to be positive.

Joint and individually estimated logit models of program participation are reported in Table 2, together with sample means, standard deviations, and expected signs. The first column of coefficient estimates provides a joint model of participation with binary shifts for hypothetical open-ended and dichotomous choice responses. Actual contribution decisions serve as the baseline. The last three columns of the table provide separate estimation results for the actual, open-ended hypothetical, and dichotomous choice hypothetical participation decisions.

In general, when significant the sign of the coefficients reflects prior expectations, and the overall models are highly significant. Favorable impressions of program characteristics tend to be positively correlated with program enrollment, although the coefficient on trees is not significant in any of the individual equations. Consistent with our expectations from the experimental provision point literature, the provision point and the money back guarantee are positively correlated with, and each is a significant explanatory variable of, participation in both the joint and most individual models. The demographic characteristics are also largely consistent with prior environmental valuation research. Participation is negatively correlated with age and positively correlated with being male or a member of an environmental organization. Neither education nor rating of service is significant in any of the regressions. Overall, the significance of each of the equations and individual explanatory variables demonstrates that responses to the questions vary in a systematic fashion.

After accounting for these covariates, the binary variables for hypothetical responses in the joint model tell a tale similar to the simple Chi-Squared test of independence. The coefficient for hypothetical open-ended responses is small and not significantly different from zero. In contrast, the dichotomous choice responses are significantly different from actual decisions at the 10 percent level.[7] These results suggest that open-ended CV provides a more accurate prediction of participation than does dichotomous choice. This result is in keeping with Brown *et al.* (1996), who find, in comparing CV to VCM results, that dichotomous choice values exceeded open-ended values, which in turn exceeded actual contributions. Our results also suggest that free-riding may explain the difference between open-ended willingness to pay and actual contributions found in the Brown

[7]Letting LL denote the log likelihood, a likelihood ratio test LR= -2($LL_{Resticted}$ - $LL_{Unrestricted}$) was used to test the null hypothesis of equality of all coefficients across equations (with the exception of a binary shift variable for each equation). The test across all three survey versions is rejected at the 5 percent level of significance (LR=40.786 > 40.113 = $\chi^2_{0.05, 27}$). Pair wise pooling of the data indicates that the null hypothesis of equality between actual and open-ended responses (LR=12.133 < 14.684 = $\chi^2_{0.05, 9}$) and actual and dichotomous choice responses (LR=12.133 < 14.684 = $\chi^2_{0.05, 9}$) cannot, however, be rejected. As such, rejection of a joint model pooling all response functions appears to be driven by the inequality of open-ended and dichotomous choice response functions (LR=35.241 > 14.684 = $\chi^2_{0.05, 9}$). Moving to paired actual-hypothetical pooled regressions, the significance of the binary shifters reflects the significance levels reported in Table 2. The coefficient on the hypothetical open ended responses was not significant at any level (t=0.48) while that on the hypothetical dichotomous choice responses was significant at the 10 percent level (t=1.91).

et al. (1996) study.

In assessing contingent valuation, Mitchell and Carson (1989) adapt the sociological concepts of criterion validity and construct validity. Criterion validity refers to the goodness of fit of CV estimates to benchmark values, such as market prices. Construct validity refers to whether CV estimates are related to explanatory variables as expected according to economic theory. Applying these measures here, open-ended responses appear to have a higher criterion validity than do dichotomous choice responses, where the reference criterion is the PP/MBG/EB mechanism used by NMPC. At the same time, the logistic analysis suggests that dichotomous choice responses perform better in terms of construct validity. That is, the dichotomous choice regression exhibits a substantially better fit (as measured by the likelihood ratio) than the open-ended response function. This suggests that open-ended responses do not vary as systematically with socio-economic characteristics as do dichotomous choice responses. However, given the relatively close correspondence of hypothetical open-ended responses with actual participation levels, these results challenge the usefulness of construct validity as a criterion for evaluating CV elicitation techniques.

4. CALIBRATION OF HYPOTHETICAL RESPONSES

As indicated in the previous section, some upward hypothetical bias remains even when an approximately demand revealing mechanism is used to provide a reference for actual willingness to pay. Two different methods have been recently suggested in the CV literature to calibrate or adjust hypothetical values obtained from open-ended or dichotomous choice responses to more closely reflect actual values. Such calibration is widely used in market research to adjust survey responses to predict actual demand. Here, we examine the accuracy of these two methods.

With respect to open-ended responses, Schulze *et al.* (1997) argue that a "disembedding" question following the open-ended question may reduce hypothetical bias by reminding respondents to only state values for the specific good in question rather than including other "embedded" values such as moral satisfaction. In the NMPC survey, this calibration was accomplished as follows. First, individuals were asked to answer an open-ended willingness to pay question as previously described. The following issue is then raised:

> *Some people say that it's hard to think about the amount you would pay for a specific program like GreenChoice, rather than for environmental programs or other good causes in general...*

and individuals are asked if their bid on the open-ended question was just for the GreenChoice™ program or if the stated WTP included values for a wider range of environmental or public causes. If the respondent indicates that, "Yes, my stated value included other causes," then they are asked to estimate the proportion of their stated value that was for the GreenChoice™ program. This "disembedded" portion is then multiplied by the original open-ended value to isolate the value for the program.

Several studies have used this approach, with self-reported embedding ranging from 20 percent (clean up local groundwater, McClelland *et al.*, 1992) to 50 percent (medium size oil spills, Rowe *et al.,* 1991). In addition to the notion that individuals are embedding their specific values within a broader stated value, two other interpretations of these self-reported adjustments have been offered. First, experimental economists have found that in repeated rounds, values tend to fall after the first bid. In other words values tend to be overstated on the first round and tend to approach induced or actual values in subsequent rounds (Davis and Holt, 1993). The disembedding question thus allows individuals to act as if they are in a more experienced, second round situation. Second, the disembedding question might act as a reminder that individuals may want to spend their money in other ways, thus providing an additional opportunity to consider budget constraints and substitutes. The need for emphasizing these constraints in the contingent valuation question was highlighted in Arrow *et al.* (1993).

Champ *et al.* (1995, 1997) have suggested an alternative debriefing method appropriate to dichotomous choice contingent valuation. Building on evidence that individual respondents have some uncertainty in their WTP values (Gregory *et al.*, 1995; Ready *et al.*, 1995; Welsh and Poe, 1998), this approach asks those who responded "Yes" to the dichotomous choice question the following debriefing question:

> *So you think that you would sign up. I'd like to know how sure you are of that. On a scale from 1 to 10, where 1 is "very uncertain" and 10 is "very certain," how sure are you that you would sign up and pay the extra $6 a month?*

Using similar wording, a CV field validity study of WTP for road removal on the north rim of the Grand Canyon (Champ *et al.*, 1997) found that estimating a model, in which only the "yes" respondents who had a certainty level of 10 were coded as yes responses, was not significantly different from actual contributions. Champ and Bishop (1998) report that a certainty level of “8 or higher” provides the best approximation of actual willingness to pay for a wind power program. As noted the Champ *et al.* (1997), and the Champ and Bishop (1998), reference for actual contributions was obtained using a voluntary contribution mechanism, which may lead to excessive calibration coefficients.

The results of these calibration approaches are reported in Table 3. On average, the open-ended respondents reported that 23 percent of their values were embedded (i.e. that their value for the program was 77 percent of their original open-ended response). This lowered the entire open-ended WTP distribution, and reduced the estimated percentage who would have said yes to $6 to 16.4 percent. Although this is lower than the actual contribution level of 20.4 percent, it (as well as the original unembedded open-ended proportion) is not significantly different from this reference value (χ^2=1.03). However, these results are suggestive in the sense that they demonstrate that overcompensation using this approach is possible. Over-correction is predicted in the psychological literature on self-correction when the subjects are given additional information on sources of error, as was done with this case in the embedding question (Wegener and Petty, 1997; Wilson and Brekke, 1994).

Dichotomous choice respondents reported a wide range of certainty levels. Only a small percent reported that they had a certainty level of 10, with the mode being at 7. Sign-up proportions accounting for different certainty level thresholds are provided in Table 3. Proportions associated with treating yes responses with subsequent certainty levels of "greater than or equal to 6" and "greater than or equal to 7" as "true" yes responses, are not significantly different from the actual sign-up rate. However, a certainty level of greater than or equal to 7 most closely corresponds to the actual sign up rate. This degree of correction is less than the Champ *et al.* (1995, 1997) and the Champ and Bishop (1998) findings, a result that is consistent with our basic argument that the VCM used in these studies under reveal actual demand. In addition to the aforementioned mechanism effects, these differences may be attributed to the good itself, or to phone versus mail formats.

A third calibration approach generally attributed to the NOAA proposed regulations (1994) suggests that WTP values be divided by two in order to correct for hypothetical bias. Using this 50 percent calibration rule on individual open-ended values resulted in an estimated participation rate of 7 percent at $6, which substantially and significantly overcorrects hypothetical responses with respect to the actual (χ^2 = 16.15). If instead, the percent of participants at $6 of the dichotomous choice hypothetical values was divided by two, then the estimated percent approximates 15 percent, which is below, but not significantly different than the actual participation rate (χ^2 = 2.59). Some caution should be taken, however, in assuming a direct correspondence between correcting DC percent and a shift in WTP. In other words, halving the percent at $6 may not directly correspond to halving the estimated WTP, due to non-linearities in underlying WTP distributions.

5. SUMMARY AND DISCUSSION

Past contingent valuation validity tests have relied on voluntary contribution mechanisms to elicit actual payments. It is well known that such a mechanism engenders free-riding and the under revelation of demand. As such, the reference point used in these field tests in inconsistent with the information needed to conduct benefit cost analyses, and likely to lead to erroneous conclusions about the degree of hypothetical bias present in the contingent valuation of public goods.

Building upon recent experimental economics research, it is argued in this paper that a provision point mechanism provides a more appropriate, approximately demand revealing reference point for contingent valuation validity tests. Using such a mechanism, the research presented here indicates that even when free-riding and awareness are controlled for, contingent valuation responses still exhibit some upward hypothetical bias. However, a lower level of calibration is needed than has previously been indicated in field research comparing actual and hypothetical contributions.

Importantly, the elicitation method used in contingent valuation matters for the reliability of the method in predicting actual participation. In particular, dichotomous choice substantially overestimates actual participation, by about 50 percent. Open-ended willingness to pay overestimates by 17 percent, but this difference was not found to be statistically significant. Finally, a comparison of the results of this field study with economics laboratory experiments that have examined hypothetical bias shows substantial parallelism. A recent survey of this literature (Schulze *et al.*, 1996) argues that, for open ended willingness to pay questions, ". . .one has to conclude that some upward bias is likely to be present." In contrast, the experimental evidence concerning dichotomous choice shows that values from dichotomous choice uniformly exceed actual values (Balistreri *et al.*, 1996; Cummings *et al.*, 1995) and also exceed those obtained using open-ended willingness to pay (Schulze *et al.*, 1996). The apparent parallelism between these laboratory studies and our results in the field suggest that laboratory research could be a powerful tool in advancing contingent valuation techniques.

Table 1: Proportion of Actual and Hypothetical Sign-Ups by Response Type

Response Type	Percent Participation (observations)
Actual Dichotomous Choice	20.4 (n = 142)
Hypothetical Open Ended (OE Hypo)	23.9 (n = 284)
Hypothetical Dichotomous Choice (DCHypo)	30.6 (n = 258)

Table 2: Estimated Logit Models by Response Category

Variable [Description]	Exp. Sign	Mean (s.d.)	Estimated Coefficients (s.e.)			
			Joint	Actual	OE Hypo	DC Hypo
Constant			-4.024 (0.856)***	-4.386 (2.184)**	-2.471 (1.167)**	-5.143 (1.602)***
D-DC Hypo [OE Hypo=1]	?	0.38 (0.49)	0.574 (0.299)*			
D-OE Hypo [DC Hypo=1]	?	0.41 (0.49)	0.142 (0.298)			
Renewables [1 to 10 scale]	+	6.38 (2.71)	0.150 (0.047)***	0.233 (0.118)**	0.119 (0.099)	0.297 (0.085)***
Trees [1 to 10 scale]	+	8.44 (2.23)	0.116 (0.065)*	0.216 (0.186)	-0.012 (0.073)	0.154 (0.116)
D-Prov. Pt. [Interest=1]	+	0.17 (0.38)	1.353 (0.259)***	1.416 (0.588)**	0.925 (0.411)**	1.868 (0.479)***
D-MBGuar [Interest=1]	+	0.48 (0.50)	0.626 (0.220)***	-0.098 (0.550)	0.734 (0.329)**	0.758 (0.425)*
Age	-	51.9 (16.3)	-0.023 (0.007)***	-0.040 (0.019)**	-0.039 (0.011)***	-0.003 (0.013)
D-Gender [Male=1]	?	0.46 (0.50)	0.463 (0.213)**	0.954 (0.517)*	0.432 (0.323)	0.224 (0.388)
D-Cgrad [Cgrad=1]	+	0.36 (0.48)	0.181 (0.223)	0.002 (0.997)	0.300 (0.321)	0.275 (0.450)
D-Enviro [Contribute=1]	+	0.24 (0.43)	1.108 (0.233)***	0.666 (0.624)	0.461 (0.346)	2.474 (0.451)***
Rate Service [1 to 10 scale]	+	8.45 (1.65)	0.074 (0.069)	0.082 (0.178)	0.154 (0.102)	-0.087 (0.134)
Chi Sq			148.93	31.10	35.99	117.11
n		620	620	128	255	237

*,**, and *** indicate 10, 5 and 1 percent significance, respecitvely.

Table 3: Actual Sign-ups and Calibrated Hypothetical Sign Ups

Response Type			Percent Participation (observations)
Actual Dichotomous Choice			20.4 (n = 142)
Hypothetical Open Ended, Revised for Embedding			16.4 (n = 280)
Hypothetical Dichotomous (DC Hypo) Choice, Revised for Certainty			
Certainty	≥	5	29.0* (n = 258)
	≥	6	24.8 (n = 258)
	≥	7	20.9 (n = 258)
	≥	8	14.0* (n = 258)
	≥	9	8.5*** (n = 258)
	=	10	6.6*** (n = 258)

*,**, and *** indicate that the proportions are significantly different from the "actual" value at the 10, 5, and 1 percent levels, respectively.

REFERENCES

Arrow, K.; Solow, R.; Leamer, E.; Portney, P.; Randner, R. and Schuman, H. "Report of the NOAA Panel on Contingent Valuation," *Federal Register*, January 15, 1993, 58(10), pp. 4601-4614.

Bagnoli, M. and Lipman, B. "Provision of Public Goods: Fully Implementing the Core through Voluntary Contributions." *Review of Economic Studies*, 1989, 56, pp. 583-601.

Bagnoli, M. and McKee, M. "Voluntary Contributions Games: Efficient Private Provision of Public Goods." *Economic Inquiry*, 1991, 29, pp. 351-366.

Balistreri, E.; McClelland, G.; Poe, G.L. and Schulze, W.D. "Can Hypothetical Questions Reveal True Values? A Laboratory Comparison of Dichotomous Choice and Open-Ended Contingent Values with Auction Values."1996, Paper presented at the ASSA meetings, San Fransico.

Baugh, K.; Brynes, B.; Jones, C. and Rahimzadeh, M. "Green Pricing: Removing the Guesswork." *Public Utilities Fortnightly*, August 1995, pp. 26-28.

Bishop, R. C. and Heberlein, T. A. "Measuring Values of Extramarket Goods: Are Indirect Measures Biased?" *American Journal of Agricultural Economics*, 1979, 61, pp. 926-30.

Bishop, R. C. and Heberlein, T. A.. "The Contingent Valuation Method," in R. L. Johnson and D. V. Johnson, eds., *Economic valuation of natural resources: Issues, theory, and applications*. Boulder, CO: Westview Press, 1990, pp. 181-204.

Brown, T.; Champ, P.; Bishop, R. and McCollum, D. "Which Response Format Reveals the Truth About Donations to a Public Good?" *Land Economics*, May 1996, 72(2), pp. 152-166.

Brynes, B.; Rahimzadeh, M.; Baugh, K. and Jones, C. "Caution Renewable Energy Fog Ahead! Shedding Light on the Marketability of Renewables." Paper presented at NARUC-DOE Conference on Renewable and Sustainable Energy Strategies in a Competitive Market, May 1995, Madison, WI.

Cadsby, C. B. And Maynes, E. "Choosing Between a Cooperative and Non-Cooperative Equilibrium: Nurses versus Economics and Business Students." 1996, Unpublished manuscript, University of Guelph.

Champ, P. A., and Bishop, R. C. “Using Contingent Donations to Predict Voluntary Provision and Benefits of a Public Good” 1998, Paper presented at the World Congress of Environmental and Resource Economists, Venice, Italy.

Champ, P. A.; Bishop, R. C.; Brown, T. C. and McCollum, D. W. "A Comparison of Contingent Values and Actual Willingness to Pay Using a Donation Provision Mechanism with Possible Implications for Calibration," in Benefits and Costs Transfer in Natural Resource Planning, *Eighth Interim Report, W-133 Western Regional Research Publication*, 1995, pp.

387-410.

Champ, P. A.; Bishop, R. C.; Brown, T. C. and McCollum, D. W. "Using Donation Mechanisms to Value Nonuse Benefits from Public Goods." *Journal of Environmental Economics and Management*, 1997, 33(2), pp. 151-162.

Conover, W. J. *Practical Non-Parametric Statistics*, 2nd Ed., 1980, John Wiley and Sons, New York.

Coursey, D. and Smith, V. L. "Experimental Tests of an Allocation Mechanism for Private, Public or Externality Goods." *Scandinavian Journal of Economics*, 1984, 86, pp. 468-484.

Cummings, R. G.; Harrison, G. W. and Rutström, E. E. "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible?" *American Economic Review* , 1995, 85(1), pp. 260-266.

Cummings, R. G. and Taylor, L. O. "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method." 1996, Working Paper, Georgia State University.

Davis, D. D. and Holt, C. A. *Experimental Economics*. 1993, Princeton: Princeton University Press, .

Dawes, R.; Orbell, J.; Simmons, R. and van de Kragt, A. "Organizing Groups for Collective Action." *American Political Science Review*, 1986, 8, pp. 1171-1185.

Dickie, M.; Fisher, A. and Gerking, S. "Market Transactions and Hypothetical Demand Data: A Comparative Study." *Journal of the American Statistical Society*, 1987, 82, pp. 69-75.

Dillman, D. A. *Mail and telephone surveys - The total design method*. 1978, New York: Wiley & Sons.

Duffield, J. W. and Patterson, D. A. "Field Testing of Existence Values: An Instream Flow Trust Fund for Montana Rivers." Paper presented at the AERE/ASSA meetings, New Orleans,1992.

Farhar, B. C. and Houston, A. H. "Willingness to Pay for Electricity from Renewable Energy." Paper presented at the 1996 ACEEE Summer Study on Energy Efficiency in Buildings, August 25-31, 1996, Pacific Grove, CA.

Gregory, R.; Lichtenstein, S.; Brown, T. C.; Peterson, G. L. and Slovic, P. "How Precise are Monetary Representations of Environmental Improvements?" *Land Economics*, November 1995, 71(4), pp. 462-473.

Groves, T. and Ledyard, J. "Optimal Allocation of Public Goods: A Solution to the 'Free Rider' Problem." *Econometrica*, 1977, 45, pp. 783-809.

Hanemann, W. M. "Welfare Evaluation in Contingent Valuation Experiments with Discrete Responses." *American Journal of Agricultural Economics*, August 1984, 66, pp. 332-341.

Harstad, R. and Marrese, M. "Behavioral Explanations of Efficient Public Good Allocations." *Journal of Public Economics*, 1982,19, pp. 367-383.

Holt, E. A. *Green Pricing Resource Guide*. The Regulatory Assistance Project, February 1997, Gardiner, ME.

Isaac, R. M.; McCue, K. F. and Plott, C. R. "Public Goods Provision in an Experimental Environment." *Journal of Public Economics*, 1985, 26, pp. 51-74.

Isaac, R. M.; Schmidtz, D. and Walker, J. "The Assurance Problem in Laboratory Markets." *Public Choice*, 1989, 62, pp. 217-236.

Isaac, R. M.; Walker, J. and Williams, A. "Group Size and The Voluntary Provision of Public Goods: Experimental Evidence Using Very Large Groups." *Journal of Public Economics*, 1994, 54, pp. 1-36.

Ledyard, J. O. "Public Goods: A Survey of Experimental Research," in J. H. Kagel and A. E. Roth, eds., *Handbook of experimental economics*. Princeton: Princeton University Press, 1995, pp. 111-194.

Marks, M. B. and Crosson, R. T. A. "Alternative Rebate Rules in the Provision of a Threshold Public Good: An Experimental Investigation." *Journal of Public Economics*, 1998, 67 pp. 195-220.

Marwell, G. and Ames, R. E. "Economists Free Ride, Does Anyone Else?   Experiments on the Provision of Public Goods, IV." *Journal of Public Economics*, 1981, 15, pp. 295-310.

McClelland, G. H.; Schulze, W. D.; Lazo, J. K.; Waldman, D. M.; Doyle, J. K.; Elliott, S. R. and Irwin, J. R. Methods for measuring non-use values: A contingent valuation study of groundwater cleanup. October 1992, Report for the U.S. Environmental Protection Agency, Contract #CR-815183,.

Mitchell, R. C. and Carson, R. T. *Using surveys to value public goods: the contingent valuation method*.1989, Resources for the Future, Washington, D. C.,.

Moskovitz, D. H. "Renewable Energy: Barriers and Opportunities: Walls and Bridges." Report for the World Resources Institute, 1992, Washington, D.C.

National Oceanic and Atmospheric Administration. "Natural Resource Damage Assessments; Proposed Rules." January 7, 1994, *Federal Register*, 59(5), pp. 1139-1184.

Navrud, S. and Veisten, K. "Using Contingent Valuation and Actual Donations to Bound the True Willingness-to-Pay." Unpublished manuscript, January 1997, Dept. Of Economics and

Social Sciences, Agricultural University of Norway.

Ready, R. C.; Whitehead, J. C. and Blomquist, G. C. "Contingent Valuation when Respondents are Ambivalent." *Journal of Environmental Economics and Management*, 1995, 29(2), pp. 181-196.

Rondeau, D.; Schulze, W. D. and Poe, G. L. "Demand Revelation in Single-Period Provision Point Mechanisms with Incomplete Information." forthcoming in *Journal of Public Economics*.

Rose, S.; Clark, J.; Poe, G. L.; Rondeau, D. and Schulze, W. D. "Field and Laboratory Tests of A Provision Point Mechanism." October 1996, Paper presented at the annual meetings of the Economic Science Association, Tucson, AZ,

Rowe, R. D.; Schulze, W. D.; Shaw, W. D.; Schenk, D. and Chestnut, L. G. *Contingent valuation of natural resource damages due to the Nestucca oil spill*. RCG/Hagler Bailly Report for the Department of Wildlife, State of Washington, British Columbia Ministry of the Environment, and Environment Canada, June 1991.

Samuelson, P. A. 1954. "The Pure Theory of Public Expenditure." *Review of Economics and Statistics*, 1954, 36, pp. 387-389.

Schulze, W. D. “Green pricing: Solution for the Potential Free Rider Problem”. Report Prepared for Niagara Mohawk Power Corporation, 1994.

Schulze, W. D.; McClelland, G.; Waldman, D. and Lazo, J. "Sources of Bias in Contingent Valuation." in D. Bjornstad and J. Kahn, eds., *The contingent valuation of environmental resources: Methodological issues and research needs*. 1996, Cheltenham, U.K.: Edward Elgar Publishing, Ltd., pp. 97-116.

Schulze, W. D.; McClelland, G.; Lazo, J. and Rowe, R. D. "Embedding and Calibration in Measuring Non-Use Values." *Resource and Energy Economics*, 1998, 20, pp. 163-178.

Seip, K. and Strand, J. "Willingness to Pay for Environmental Goods in Norway: A Contingent Valuation Study with Real Payment." *Environmental and Resource Economics*, 1992, 2, pp. 91-106.

Smith, V. L. "An Experimental Comparison of Three Public Good Decision Mechanisms." *Scandinavian Journal of Economics*, 1979, 81, pp. 198-215.

Smith, V. L. "Experiments with a Decentralized Mechanism for Public Goods Decision." *American Economic Review*, 1980, 70, pp. 584-599.

Suleiman, R. and Rapoport, A. "Provision of Step-Level Public Goods with Discontinuous Contribution." *Journal of Behavioral Decision Making*, 1992, 5, pp. 133-153.

Wegener, D. T. and Petty, R. E. "The Flexible Correction Model: The Role of Naive Theories of

Bias in Bias Correction." *Advances in Experimental Social Psychology*, 1997, 29, pp. 141-208.

Welsh, M. P.; and Poe, G. L. “Elicitation Effects in Contingent Valuation: Comparisons to a Multiple Bounded Discrete Choice Approach”, forthcoming in *Journal of Environmental Economics and Management*, 1998, 36(2).

Wilson, T. D. and Brekke, N. "Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations." *Psychological Bulletin*, 1994, 116(1), pp. 117-142.

Wood, L. L.; Desvousges, W.H.; Kenyon, A. E.; Bala, M. V.; Johnson, F. R.; Iachan, R. and Fries, E. E. "Evaluating the Market for "Green" Products: WTP Results and Market Penetration Forecasts." April 1994, Working Paper #4, Center for Economics Research, Research Triangle Institute.

CHAPTER 5
A Comparison of Hypothetical Phone and Mail Contingent Valuation Responses for Green Pricing Electricity Programs*

1. Introduction

An unresolved issue in contingent valuation (CV) is whether personal and phone interviews provide more valid or “better” responses than mail survey techniques. Mail surveys presently dominate CV research because they are easier to implement and cost substantially less than phone interviews, thus enabling a greater amount of survey research to be conducted. However, in an influential review, the National Oceanographic and Atmospheric Administration (NOAA) Panel on contingent valuation maintained that it is “unlikely that reliable estimates of values could be elicited with mail surveys,” and recommended in-person or phone interviews [Arrow *et al.* 1993, p. 4611]. In response to the Panel’s report, Dillman (1993) argued that problems with personal interviews are generally understated, while the shortcomings of mail surveys are overstated. Others like Bishop *et al.* (1988: p. 336) found that “the data gathered through self-administered surveys may, other things being equal, be better than that obtained by telephone surveys.”

Evidence on survey elicitation “mode” effects is drawn primarily from work by survey researchers in fields other than CV. Unfortunately, these studies provide little evidence about how willingness to pay (WTP) is affected by survey mode. Two exceptions are found in recent CV phone/mail mode comparisons performed by Whittaker *et al.* (1998) and by Loomis and King (1994). Loomis and King elicited WTP for improved wildlife habitat in California. However, they attained relatively low response rates for both the mail and phone portions of their survey and drew samples from slightly different populations, which reduces the comparability of CV responses by increasing the likelihood of differing sample frames. Whittaker *et al.*elicited WTP for increased fees

***Authors: Robert G. Ethier, Gregory L. Poe, William D. Schulze, and Jeremy Clark**

at state parks, achieving high response rates with a survey from a self-selected sample of Colorado State Park users. Yet in assessing their work, they note that fee increases involve strategic issues and that their sample was not drawn from the general population. A further limitation of these two studies, from the perspective of CV validity, is that neither had actual behavior to serve as a reference for evaluating mode effects. Thus, the critical question of whether phone or mail responses provide better estimates of actual behavior remains unanswered by existing CV research.

This paper offers a unique contribution to the CV survey literature by providing a direct telephone/mail comparison with response rates in the 70% range (as suggested by the NOAA Panel), utilizing a sample drawn from a general population (a sampling strategy corresponding to most CV research), and eliciting WTP for a good with high non-use value (a common CV application). In addition, we are able to compare hypothetical phone and mail results with actual commitments to a Niagara Mohawk Power Corporation green electricity pricing program. Calibration of responses across modes, as suggested by Champ *et al.* (1997), is also explored to investigate whether different modes necessitate different levels of correction.

The remainder of the paper is organized as follows. Potential mode related survey errors and biases are briefly reviewed in the next section. Section 3 reviews the existing CV research on mode effects for the purpose of formulating a set of testable research hypotheses. The fourth section summarizes the survey methodology. Results and discussion are provided in Sections 5.

2. SURVEY ERRORS AND POSSIBLE MODE EFFECTS

Past research has demonstrated that CV responses are characterized by *hypothetical bias,* in that "hypothetical questions, particularly about donations to generally desirable environmental goods, seem to engender overestimates of actual WTP." [Brown *et al*. 1996, p. 164]. A separate body of research suggests that there are mode effects on survey responses, including stated WTP. Yet, because these two concepts have not been linked in a single study, it remains to be determined whether phone or mail elicitation acts to minimize hypothetical bias in CV research, and whether different types of calibration rules are needed across modes.

In addition to this broad concern about hypothetical bias, the survey nature of CV introduces other possible errors in drawing inferences about social values for public goods. *Sampling error* is the first source of error likely to be encountered in the survey process. This error may be induced by differences in how the sample is drawn, resulting in a divergence between the survey and target populations if elements of the target population are systematically excluded from the sample frame. Thus if alternative modes rely on different sample frames, and the sample composition affects estimates of sample values, then mode effects may be observed. Such divergence has been termed sample frame bias, coverage bias, or non-coverage error [Edwards and Anderson, 1987; Loomis and King; Groves, 1987; Dillman, 1991]. A standard example of this type of bias is that telephone directory listings often used in survey research exclude a substantial portion of the population (perhaps up to 35 to 40 percent or higher) because of unlisted numbers, recent moves, and non-phone households [Schuman, 1996; Traugott *et al.*, 1987].

Another form of sampling error that may occur is *non-response bias*. If the decision to complete and return the survey is systematically related to individual attributes, the resulting sample may not accurately reflect the population being sampled. Inferences about population values may be biased because a non-representative sample would result. If alternative survey modes have different patterns of non-response, then alternative modes would be expected to exhibit different degrees of this type of bias. Quantification and subsequent weighting of responses to account for sampling effects have been suggested as a technique to correct for non-response [Loomis, 1987;

Dalecki *et al.*, 1993]. However, the extent to which respondents' answers would differ from those of non-respondents is difficult to quantify and is seldom known. One solution is to keep the proportion of non-respondents as low as possible [Schuman, 1996], with the NOAA directives identifying 70% as an appropriate response rate for reducing the likelihood of non-response bias to acceptable levels. At this or similar response rates, "relatively few (mode) differences have been reported for personal attributes", suggesting that reported demographics can be treated as an accurate representation of the sample population [Dillman, 1991, pp. 241-42].

Suppose that appropriate sampling methods and high response rates generate comparable mail and phone samples when judged by observable "objective" demographic characteristics. It is still possible that different modes will not produce the same answers to non-demographic questions. This could occur because of *sample selection bias*, where respondents having the same observable demographic characteristics, but different unobservable characteristics, respond to a survey with a different likelihood. Such sample selection bias might be attributed to an unobservable *avidity effect*, where households with high interest in an issue are more likely to respond to a survey about that issue even though their observable demographic characteristics do not differ from the rest of the population [Loomis and King; Mitchell and Carson, 1989]. Arguments made by Arrow *et al.* and others suggest that avidity effects are more likely to occur in mail surveys because of the ease of non-response (by the non-avid) relative to telephone surveys. As with non-response bias, high response rates are offered as protection against sample selection bias [Dillman, 1991].

A third factor that might cause a deviation between phone and mail responses is *measurement error*. Differential measurement errors would result in the same respondent providing different answers to the same questions across survey modes. Schuman and others [Schuman and Presser, 1981; Dillman, 1991] have noted that phone respondents tend to give more extreme answers to survey questions. A prime explanation for this, especially for relatively uncomplicated survey topics, is that some respondents may desire to avoid embarrassment and to project a favorable image to the researchers.[1] Such a *social desirability bias* would be expected to affect, in predictable directions, a wide range of questions of a "subjective" nature which might reflect upon the respondent. For example, Connelly and Brown (1994) showed that check-off contributions to wildlife on New York State tax forms were over reported in a mail survey by more than twice actual contributions after adjusting for recall errors. Katosh and Traugott (1981) found that reported voting behavior overstated support for the winning candidate by over 10 percent. Moreover, the broad survey literature suggests that social desirability bias will be higher for both phone and in-person interviews than for mail surveys [Whittaker *et al.*; Dillman, 1978]. As an example, Dillman and Tarnai (1991) found that the frequency of socially desirable answers given in phone surveys is significantly higher than those in mail surveys for questions about driving while intoxicated. For CV, evidence of social desirability effects on WTP responses has been conjectured for phone and in-person surveys [Arrow *et al.*]. However, this hypothesis has only been supported indirectly by evidence that WTP is affected by different interviewers in in-person interviews [Boyle and Bishop, 1988; Mannesto and Loomis, 1991].

3. PREVIOUS PHONE/MAIL CV COMPARISONS AND TESTABLE PROPOSITIONS

To our knowledge only two CV studies have explicitly compared phone and mail implementations of the same survey. Loomis and King compared phone and mail WTP for the improvement of wildlife habitat and fisheries resources in California. However, their sampling frames differed substantially: they used random digit dialing for the phone survey, but supplemented phone listings with auto registrations for the mail survey. Statistically significant differences between mail and phone respondents were found for four demographic characteristics: gender, education level, age and income. This result is not particularly surprising in light of their low response rates (35% mail and 55% phone), suggesting a combination of sample non-response and sample-selection bias. In spite of these differences in underlying demographics, Loomis and King did not find that broad attitudes about wildlife differed between survey modes. Nevertheless, mail surveys produced higher WTP estimates for four of the five hypothetical programs. In one case, the estimated mail WTP value was more than triple the estimated phone value. Loomis and King attributed higher mail WTP values to “sample selection effects in mail surveys” (p.318), which is consistent with the sample non-response bias and avidity effects discussed above.[2]

The study by Whittaker *et al.* used a split sample design with identical phone and mail instruments to ask Colorado state park users about their WTP for increased user fees to be paid by all park users at Colorado state parks. Drawing from a homogeneous sample across modes, the survey produced high response rates (78% mail, 79% phone), thus minimizing the possibility of sampling error. Such high response rates might be anticipated because their survey population consisted of state park users who had previously completed a one-page survey and who volunteered their addresses and phone numbers. Statistical tests were conducted to evaluate differences between modes for four variables (number of years visiting Colorado state parks, number of parks visited per year, expenditures on recreation equipment, and household income). Only income was found to be significantly different across phone and mail surveys. Since a limited number of dichotomous choice fee values were used, a WTP distribution was not estimated. Instead, using Chi-Square tests for each fee level, Whittaker *et al.* found that phone respondents were more likely than mail respondents to say yes to a fee increase, in one case by a 67% to 43% margin. These findings are opposite those of Loomis and King. Whittaker *et al.* attribute this reversal in part to social desirability bias created by the presence of an interviewer, which is likely to be pronounced because of the sample used and the commodity selected. That is, the conditions were almost ideally designed to induce social desirability effects: respondents were obviously interested in the commodity and, because they had already been contacted in the park, were more likely to perceive that interviewers would be favorable to resource conservation. Whittaker *et al*. also suggest that there may have been strategic bias, in that respondents may seek to discourage the park agency from raising fees. They argued that strategic bias would comparatively lower mail results because respondents would have more time to develop strategic thought.

Taken together, the research literature on mail and phone surveys in other disciplines, as well as the specific CV work by Loomis and King and Whittaker *et al*., suggests that there may be important, systematic differences between responses obtained from phone and mail survey techniques. However, the likelihood of sampling errors in the Loomis and King study, and the possible self-selection and strategic effects in the Whittaker *et al*. study, limit the relevance of these studies to broader CV research. Using samples with identical coverage, drawn from general populations, our research further explores some of these issues by examining five propositions. Unless otherwise noted, non-parametric r x c Contingency Table analyses are used to examine each proposition [Conover, 1980].

<u>Proposition 1:</u> *Phone and mail surveys do not differ in sample response bias as evidenced by observable respondent demographics.* If distributions of sample demographics are not determined to be significantly different between phone and mail surveys, we conclude that sample and non-response bias does not differ across modes. Such a result would allow us to attribute any observed mode differences in WTP to alternate explanations (i.e., avidity, social desirability, hypothetical bias).

<u>Proposition 2:</u> *Phone and mail surveys do not yield different responses to non-WTP questions.* As discussed, avidity effects are more likely to occur in mail surveys because of the ease of non-response, while social desirability bias is thought to be more prevalent in a phone survey where the presence of an interviewer may influence respondent answers. While we are unable to separate avidity and social desirability effects in this research, we can compare responses to non-WTP questions to gauge the relative effects of such biases. If the green pricing program is thought to be a socially responsible program, avidity and social desirability bias would work in the same direction but on different survey modes. Existence of such effects is evaluated by comparing response patterns to four questions likely to be inflated by either bias. If the patterns do differ, a larger mean response for the mail survey would support the conclusion that avidity effects outweigh any social desirability effect, while a larger phone survey mean value would suggest that social desirability bias dominates.

<u>Proposition 3:</u> *Phone and mail surveys do not lead to different hypothetical participation decisions.* Whittaker *et al.* concluded that the use of phone surveys results in higher willingness to say ‘yes’ to a hypothetical fee increase for all park users, which might be caused by social desirability bias not present in a mail survey. In contrast, Loomis and King found that mail surveys exhibited higher WTP than phone elicitations. Given these conflicting results, the null hypothesis is that hypothetical phone and mail participation results are the same.

<u>Proposition 4:</u> *Phone and mail surveys do not differ in the direction and degree of hypothetical bias.* Even if mode effects do occur, it is not apparent which method provides more accurate predictions of actual participation levels. Alternative methods might systematically overstate, understate, or bracket “true” WTP. Here, we compare hypothetical and actual “yes” responses to dichotomous choice questions of willingness to participate in a green electricity pricing program. If actual commitments differ from hypothetical sign-ups, we conclude that there is hypothetical bias.

Combining these results with those from Proposition 3 allows us to draw conclusions about relative hypothetical bias across modes.

Proposition 5: *Phone and mail participation estimates demonstrating hypothetical bias may be corrected with identical calibration*. If hypothetical bias is found to exist in both survey modes, an interesting and useful question is whether similar calibration methods provide similar participation estimates. Following Champ *et al.* (1995, 1997), we included a calibration question immediately following the dichotomous choice WTP question. In this approach, participants who respond "yes" to the hypothetical participation question are asked how certain (on a scale from 1 to 10, where 1 is "Very uncertain" and 10 is "Very certain"), they are that they would sign up for the program if it were indeed offered. Calibrated participation rates based on these certainty responses are compared.

4. SURVEY METHODOLOGY

In 1995-1996 the Niagara Mohawk Power Corporation (NMPC), a public utility in New York State, launched Green Choice™, the largest program in the country for the green pricing of electricity [Holt and Associates, 1997]. NMPC's 1.4 million households were offered the opportunity to fund a green electricity program that would invest in renewable energy projects (e.g., landfill gas reclamation, wind power) as substitutes for traditional energy sources, and a tree planting program. Such green pricing programs have generated substantial interest as utilities come under increasing pressure to provide alternative sources of electricity for customers who prefer environmentally friendly energy sources. Although windpower stations in Traverse City, Michigan, and Ft. Collins, Colorado, have been funded through voluntary surcharges on electricity bills, comparisons of market research and actual sign-up rates for a range of green pricing programs generally suggest that the high levels of interest predicted from marketing research have been seriously biased. While marketing research suggests that 'majorities' of customers will sign-up for such programs, sign up rates have generally been below two percent [Brynes *et al.*, 1995; Holt, 1997].

Building on the mechanism design recommended by Schulze (1994), NMPC's Green Choice™ provision mechanism incorporated three key features: a provision point, a money back guarantee, and extended benefits. This provision point mechanism works as follows. If, for example, 12,000 customers joined the Green Choice™ program in the first year at a fixed $6/month (collected as a surcharge on their electricity bill), $864,000 would be collected and the provision point would be met. NMPC would then plant 50,000 trees and fund a landfill gas project which could replace fossil fuel generated electricity for 1,200 homes. However, if Green Choice™ participation was insufficient to reach the provision point, NMPC would cancel the program and refund all the money that was collected. If money were collected in excess of the provision point, additional funds would be applied toward increasing the scope of the program. The characteristics of the program itself were based on prior market research for NMPC [Wood *et al.*, 1994]. The demand revelation characteristics of the program's funding mechanism are further discussed in an experimental context in Rose *et al.* (1997).

Provision point, money back guarantee mechanisms have been shown to improve participation levels in public programs [Bagnoli and McKee 1991; Schulze], and to approximate aggregate demand in 'large group' laboratory experiments [Rondeau *et al*. 1998]. Having a demand revealing mechanism is an important component of validity testing. As argued in Poe *et al.* (1997), Navrud and Veisten (1997), and Foster *et al.* (1997), previous public goods CV validity tests that compared hypothetical and "simulated" markets (e.g., Duffield and Patterson, 1992; Seip and Strand, 1992; Brown *et al.*; and Champ *et al.* 1997) have used voluntary contribution mechanisms which have been demonstrated to under reveal 'true' demand, thus providing biased estimates of hypothetical bias. Because the provision point mechanism has been demonstrated to more closely approximate demand by reducing free-riding, it should better reflect actual preferences. In turn, this should allow for more accurate determination of hypothetical bias and provide an appropriate reference for calibration.

The survey instruments used in this study followed the Dillman Total Design Method [Dillman, 1978]. The survey was pretested by administering successive draft versions by phone until respondents clearly understood the instrument. A split-sample was used to compare three designs: actual sign-ups obtained through phone solicitation, and phone and mail responses to hypothetical versions of the actual solicitation question. Initial sample sizes were 250 households for actual sign-ups, and 500 households each for hypothetical phone and mail surveys. Established multiple contact survey techniques, including a two dollar incentive, were used in all versions with Cornell University as the primary correspondent. All three survey versions were administered by Hagler-Bailly Inc. The subject population was NMPC customers in the greater Buffalo, New York, metropolitan area. Sample frames were obtained from Genesys, Inc., a private marketing firm, which used identical methods to draw the samples of households' addresses with phone numbers.

In each survey version, respondents were first screened to establish that they were NMPC customers and to determine their previous knowledge of the GreenChoice™ program. A description of the GreenChoice™ program followed, with questions to aid the respondents' understanding. The program description corresponded with the NMPC Green Choice™ brochure as closely as possible and emphasized various components of the good (trees and renewable energy) and the provision point mechanism. The description was followed by a dichotomous choice question asking the respondents if they would sign up for the program. The survey concluded with demographic questions.

Adopting the dichotomous choice calibration method suggested in Champ *et al.* (1997), participants in the hypothetical surveys were first asked if they would be willing to participate in the program if it was offered at $6/month, and, if they answered "yes", were next asked how certain they were of actually participating in the program on a scale from '1' ("Very Uncertain") to '10' ("Very Certain"). Participants in the actual phone survey were offered the opportunity to sign up for the program at $6/month, with the charge to appear on their monthly bill. This sign-up now/pay later approach follows standard green pricing methods [Holt]. Furthermore, the phone solicitation approach corresponded with the "keep it simple" approach adopted by NMPC, which allowed such phone sign-ups for the actual program.[3]

One of the motivations for using mail surveys is their relatively low cost per respondent compared to a phone survey.[4] In this study the final cost per respondent varied by nearly a factor of two between the phone and mail treatments: the final cost per telephone respondent was estimated to be $45.95, while the final cost per mail respondent was estimated to be $24.12. Thus, all else constant, if phone surveys were found to provide superior valuation results, and thus became the preferred valuation mode, it is likely that fewer CV studies would be conducted and that those studies would have smaller sample sizes. In contrast, if mail and phone surveys are found to produce comparable results, less expensive mail survey techniques could continue to be used.

4. SURVEY RESULTS

Table 1 provides information on response and participation rates by survey mode. The adjusted sample size proportions (after excluding undeliverables, disconnected phones, non-NMPC customers, those who had previously heard of the Green Choice™ program, deceased, and individuals not responsible for the electricity bill) are significantly different across the three samples

($\chi^2 = 9.75 > 9.21 = \chi^2_{2,0.01}$) indicating statistically significant sampling effects. Thus, even though the initial sample is drawn using identical sampling techniques, the different contact methods appear to induce different levels of attrition. One should note that, to a large extent, this deviation is driven by screening questions unique to this survey (e.g., “Are you a Niagara Mohawk customer?”, “Have you heard or read anything about the Green Choice program?”) as the proportion of “List Errors” are not significantly different between the three samples ($\chi^2 = 2.00 < 4.61 = \chi^2_{2,0.10}$) while the “Screening” proportions are significantly different ($\chi^2 = 7.74 > 5.99 = \chi^2_{2,0.05}$). Adjusted response rates as measured by the number of completed surveys suggest that non-response effects, if they exist, do not differ significantly across the three samples ($\chi^2 = 1.66 < 9.21 = \chi^2_{2,0.05}$).

We now turn to statistical evaluation of the propositions.

Proposition 1: *Phone and mail surveys do not differ in sample response bias as evidenced in observable respondent demographics.* Consistent with the statistically similar response rates, evidence of response effects is not found in the demographic variables. Using r x c Contingency Table analyses the distributions of five demographic variables were compared across hypothetical phone and mail surveys: age, gender, income, occupation, and level of schooling. None was significantly different at the 10% level of certainty (see Table 2). [5] Thus, there is no observable difference in sample demographics, suggesting similar samples. This result contrasts with both the Loomis and King and Whittaker *et al.* studies, which found differences in their respondent demographics. Because there is no evidence of differences in sample demographics across modes in our survey, any differences in survey responses across modes found while examining the subsequent propositions will be attributed to other mode effects and not to sample effects or non-response bias.

<u>Proposition 2:</u> *Phone and mail surveys do not yield different responses to non-WTP questions.* Four questions which might be expected to capture both an avidity effect or social desirability bias were asked: “In the last two years, have you contributed to any environmental groups…?”, “Please rate the overall service you receive from NMPC?”, “How interested are you in renewable energy?”, and “How interested are you in the goal of planting trees on public lands in Upstate New York?”. Chi-square r x c Contingency Table test results for the four questions are shown in Table 3. We found differences in three of the four “subjective” questions asked in each survey. Differences in ‘Rate Service’ and ‘Planting trees’ are significant at the 1% level, while ‘Give to environment’ is significant at the 5% level. Interest in “Renewable Energy” is not significantly different across modes. For all four questions evaluated, the phone survey produced higher mean responses than the mail survey. This suggests that social desirability bias present in the phone survey may outweigh any avidity effects which might have been captured in the mail survey. Such a finding is consistent with Dillman (1991), as well as with the conclusions of Whittaker *et al.*, but contrasts with Loomis and King, who may have had an artificially high avidity effect associated with sample non-response bias.

<u>Proposition 3:</u> *Phone and mail surveys do not lead to different hypothetical participation decisions.* The results of this hypothesis are subject to interpretation. On one hand the hypothetical mail (35.5%) sign-up rate was about 17 percent higher than the hypothetical phone (30.6%) sign-up rate, a difference that is arguably large. Yet, unlike previous studies, this difference is not significantly different at the 10% level ($\chi^2 = 1.44 < 2.71 = \chi^2_{1,\ 0.10}$), even though the sample size used in these comparisons exceeds the single bid value samples typically used in dichotomous choice CV studies.[6] Note also that the direction of the observed difference between the hypothetical results is the opposite of what one would expect if the phone answers were influenced by social desirability bias, as was found for Proposition 2. Given this divergence, we speculate that the nature of the WTP

question, with the inclusion of a dollar amount, works to reduce the effect of the social desirability bias found in Proposition 2 by 'grounding' the respondent in a way that may not occur in an 'opinion' question. Thus, social desirability bias might be present in non-WTP questions in a manner consistent with previous research in other disciplines, but without carryover to valuation questions.

These results contrast with Whittaker *et al*. but are consistent with Loomis and King. One means of reconciling our results with those of Whittaker *et al.* is to emphasize their secondary explanation for low mail WTP levels: strategic respondents trying to avoid the possibility of an actual fee increase. Since our hypothetical WTP question involves only the voluntary, fixed price payment by the respondent, strategic incentives to influence actual policies by under-revealing WTP are limited [Hoehn and Randall, 1987]. Indeed, the reverse strategy might be possible: respondents might seek to convince NMPC that green power is highly valued and thus ought to be provided. In addition, the on-site selection of the sample from park visitors used in Whittaker *et al.* may be more likely to exhibit social desirability bias than a sample drawn from a more general population. Regardless of cause, the WTP results are not found to be significantly different and we conclude that hypothetical phone and mail WTP results are comparable.

Proposition 4: *Phone and mail surveys differ in the direction and degree of hypothetical bias.* The proportion of respondents to the actual survey who chose to join the Green Choice™ program at $6/month was 20.4%. The comparable proportions for the hypothetical surveys were 30.6% for the phone sample and 35.5% for the mail survey. These results are summarized in Table 4.

In assessing these relative sign-up rates, it is important to note that *a priori* we expected actual participation rates in this research to be substantially higher than the two percent (or less) sign-up rates observed for other green electricity programs. In part this divergence may be attributed to the provision point mechanism used to elicit participation in this field experiment. But the most important factor is likely to be differences in program awareness in this study as compared to previous green pricing program implementations. The participation rate in other green programs has typically been defined in terms of the total customer base or the proportion of the customer base that received direct mailings. Our research instead defines response rates in terms of the number of completed phone and mail surveys, thus ensuring 100% awareness. Participation rate estimates from marketing studies have also not been adjusted for awareness, which is expected to be low for typical utility marketing programs [Farhar and Houston, 1996; Holt].

Contingency table analyses show that the actual sign-up rate of 20.4% was indeed significantly different from either of the hypothetical sign-ups. Both the actual-hypothetical phone difference ($\chi^2 = 4.83$) and the actual-hypothetical mail difference ($\chi^2 = 10.09$) are significant at the 10% level ($\chi^2_{1, 0.10} = 2.71$). Both hypothetical questions overestimate actual contributions. Therefore, if hypothetical dichotomous choice responses are to be used to accurately predict actual enrollment, some sort of calibration will be necessary even when awareness is controlled and a demand revealing mechanism is used to collect actual commitments.

<u>Proposition 5:</u> *Phone and mail participation estimates demonstrating hypothetical bias may be corrected with identical calibration.* While differences in hypothetical phone and mail WTP responses may be subject to interpretation, the calibration of these hypothetical responses produced remarkably similar phone and mail WTP patterns (see Table 4). None of the hypothetical phone-mail differences across calibrated values is statistically significant. This result is important. It suggests that while mode effects might affect WTP responses, calibration of these responses produces similar valuation patterns. Since the objective is to predict “true” WTP, it appears that the alternative modes do not provide different estimates of WTP once they are calibrated.

In this analysis, calibration levels of ‘7' or greater for each mode were nearly identical to the actual sign-up rate of 20.4%: hypothetical phone sign-ups at a certainty level of ‘7' or greater were 20.9%, while mail sign-ups at a certainty level greater than or equal to ‘7' were 22.0%. This convergence suggests that a certainty level of seven or greater is an effective way of approximating actual WTP from a hypothetical survey, although calibrations of ‘6' and ‘8' (mail only) are also not found to be significantly different from actual participation levels. It is important to note that these results differ from those of Champ *et al*.[1997] and Champ and Bishop [1998], who found that certainty levels of ‘10’ and ‘8' or greater, respectively, corresponded to actual program enrollments. It is possible that calibration levels may vary with commodity. We argue that a likely explanation for the difference between calibration levels is our use of a demand revealing mechanism for the actual commitments.[9] Recall that the voluntary contribution mechanism used to collect actual payments in the Champ *et al*. studies is expected to under-reveal true WTP, and thus may lead to over-calibration through use of a higher level of certainty. Regardless of which calibration level is ultimately judged to be most appropriate, however, the pattern of calibrated levels is nearly identical across modes.

5. SUMMARY

With the increased application of CV there has been substantial debate over whether CV surveys should be conducted by mail, phone or in-person interviews. Unfortunately most of the arguments for and against different models largely rely on experiences from survey research conducted in other disciplines. To date, little empirical CV research on the relationship between mode and WTP has been conducted. The few studies that have attempted to address this issue in CV research provide results that may have limited relevance to broad CV research because of their specific design features.

In contrast to previous CV efforts to examine mode effects on survey response, this survey achieved high response rates for an environmental good with a high nonuse component, and used identical sampling techniques to draw samples from a general population. This is also the first study of mode effects to compare different hypothetical methods to an actual willingness to participate.

Using a green electricity pricing program as a case study, a comparison of hypothetical mail and phone sign-up rates found no difference in demographic variables when survey response rates approximated the NOAA threshold of 70%. Thus, measurable sampling and non-response bias does not appear to be a problem when response rates approach accepted levels. In spite of the similar demographics, mode effects *are* observed in responses to subjective, non-WTP ‘opinion’ questions. Notably, social desirability bias seems to affect phone answers, resulting in higher responses to what might be regarded as socially desirable ‘opinion’ questions. However, such social desirability effects attributed to phone elicitation do not appear to carry over to the WTP question. Indeed, the proportion of hypothetical sign-ups was found to be lower by phone than by mail, but not significantly so.

Consistent with previous CV validity research, hypothetical bias was found for dichotomous choice participation questions, with both hypothetical mail and phone survey responses significantly overstating actual commitments obtained through a phone sign-up program. An important finding of this research is that calibration questions based on certainty of “yes” responses provided statistically identical patterns across modes. This finding suggests that if calibration techniques are to be used to adjust hypothetical dichotomous choice responses to more closely reflect actual WTP, then the calibrated responses will be independent of survey mode.

In all, some evidence of social desirability and mode effects does appear across modes, but these effects do not carry over to phone and mail estimates of willingness to participate, whether calibrated or uncalibrated. As such, neither mode appears to dominate from the perspective of providing more valid estimates of actual participation decisions. We therefore argue that the selection of survey mode must be based on other criteria.

References

Arrow, K., R. Solow, E. Leamer, P. Portney, R. Randner, and H. Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation." Federal Register 58 (10): 4601-4614.

Bagnoli, M., and M. McKee. 1991. "Voluntary Contributions Games: Efficient Private Provision of Public Goods." Economic Inquiry 29:351-366.

Balistreri, E., G. McClelland, G. Poe, and W. Schulze. 1996. "Can Hypothetical Questions Reveal True Values? A Laboratory Comparison of Dichotomous Choice and Open Ended Contingent Valuation with Auction Values." Department of Agricultural, Resource, and Managerial Economics, Working Paper Series in Environmental and Resource Economics 96-01, Cornell University.

Bishop, G., H. Hippler, N. Schwarz, and F. Strack. 1988. "A Comparison of Response Effects in Self-Administered and Telephone Surveys." In Telephone Survey Methodology, ed. R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, and J. Waksberg. New York: John Wiley & Sons.

Boyle, K. J., and R. C. Bishop. 1988. "Welfare Measurement Comparisons." American Journal of Agricultural Economics 70 (1):20-28.

Brown, T. C., P. A. Champ, R. C. Bishop, and D. W. McCollum. 1996. "Which Response Format Reveals the Truth About Donations to a Public Good?" Land Economics 72 (May):152-166.

Brynes, B., M. Rahimzadeh, K. Baugh, and C. Jones. 1995. "Caution Renewable Energy Fog Ahead! Shedding Light on the Marketability of Renewables." Paper presented at the NARUC-DOE Conference on Renewable and Sustainable Energy Strategies in a Competitive Market, Madison, WI. May 1995.

Champ, P.A., R.C. Bishop, T.C. Brown, and D.W. McCollum. 1995. "A Comparison of Contingent Values and Actual Willingness to Pay Using a Donation Provision Mechanism with Possible Implications for Calibration." In Benefits and Costs Transfer in Natural Resource Planning, Western Regional Research Publication W-133, 8:387-410.

-----. 1997. "Using Donation Mechanisms to Value Nonuse Benefits from Public Goods." Journal of Environmental Economics and Management 33 (June):151-162.

Champ, P. A, and R. C. Bishop. 1998. "Using Contingent Donations to Predict Voluntary Provision and Benefits of a Public Good." Paper presented at the World Congress of Environmental and Resource Economists, Venice, Italy.

Connelly, N. A. and T. L. Brown. 1994. "Effect of Social Desirability Bias and Memory Recall on Reported Contributions to a Wildlife Income Tax Checkoff Program." Leisure Sciences 16: 81-91.

Conover, W. J. 1980. Practical Non-Parametric Statistics, 2nd Edition. New York: John Wiley & Sons.

Cummings, R. G., G. W. Harrison, and E.E. Rutström. 1995. "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible?" American Economic Review 85 (1): 260-266.

Dalecki, M. G. T., J. C. Whitehead, and G. C. Blomquist. 1993. "Sample Non-Response Bias and Aggregate Benefits in Contingent Valuation: An Examination of Early, Late, and Non-Respondents." Journal of Environmental Management 38:133-143.

Dillman, D.A. 1978. Mail and Telephone Surveys: The Total Design Method. New York: John Wiley & Sons.

----. 1991. "The Design and Administration of Mail Surveys." In Annual Review of Sociology, ed. W.R. Scott, and J. Blake. Palo Alto, CA.

----. 1996. Letter to the Office of Policy, Planning and Evaluation, U.S. EPA, 1993, reprinted in The Contingent Valuation of Environmental Resources, ed. D. Bjornstad, and J.R. Kahn. Brookfield, VT: Edward Elgar.

Dillman, D.A., and J. Tarnai. 1991. "Mode Effects of Cognitively Designed Recall Questions: A Comparison of Answers to Telephone and Mail Surveys." In Measurement Errors in Surveys, ed. P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman. New York: John Wiley & Sons.

Duffield, J. W., and D.A. Patterson. 1992. "Field Testing of Existence Values: An Instream Flow Trust Fund for Montana Rivers." Paper presented at the AERE/ASSA meetings, New Orleans, January 1992.

Edwards, S., and G. Anderson. 1987. "Overlooked Biases in Contingent Valuation Surveys." Land Economics 63 (2):168-178.

Farhar, B. C., and A. H. Houston. 1996. "Willingness to Pay for Electricity from Renewable Energy". Paper presented at the 1996 ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA, Aug. 25-31.

Foster, V., I. J. Bateman, and D. Harley. 1997. "Real and Hypothetical Willingness to Pay for Environmental Preservation: A Non-Experimental Comparison." Journal of Agricultural Economics 48 (2):123-138.

Groves, R. M. 1987. "Research on Survey Data Quality." Public Opinion Quarterly 51:156-172.

Hoehn, J. P., and A. Randall. 1987. "A Satisfactory Benefit-Cost Estimator from Contingent Valuation." Journal of Environmental Economics and Management 14:226-247.

Holt, E. A. 1997. Green Pricing Resource Guide. Gardiner, ME: The Regulatory Assistance Project.

Katosh, J. P., and M. W. Traugoff. 1981. "The Consequences of Validated and Self-Reported Voting Measures." Public Opinion Quarterly 45:519-535.

Loomis, J. 1987. "Expanding Contingent Value Sample Estimates to Aggregate Benefit Estimates: Current Practices and Proposed Solutions." Land Economics 63:396-402.

---- and M. King. 1994. "Comparison of Mail and Telephone-Mail Contingent Valuation Surveys." Journal of Environmental Management 41:309-324.

Mannesto, G., and J. Loomis. 1991. "Evaluation of Mail and In-Person Contingent Valuation Surveys: Results from a Study of Recreational Boaters." Journal of Environmental Management 32:177-190.

Mitchell, R., and R. Carson. 1989. Using Surveys to Value Public Goods: The Contingent Valuation Method. Washington, D.C.: Resources for the Future.

Navrud, S., and K. Veisten. 1997. "Using Contingent Valuation and Actual Donations to Bound the True Willingness-to-Pay." Department of Economics and Social Sciences and Department of Forestry, unpublished manuscript, Agricultural University of Norway.

Poe, G. L., J. Clark, and W. D. Schulze. 1997. "Can Hypothetical Questions Predict Actual Participation in Public Programs: A Field Validity Test Using a Provision Point Mechanism." Department of Agricultural, Resource, and Managerial Economics, Working Paper Series in Environmental and Resource Economics 97-05 , Cornell University. Paper presented at the European Association of Environmental and Resource Economists, Tillburg, June.

Rondeau, D., G.L. Poe, and W.D. Schulze. 1997. "Developing a Demand Revealing Market Criterion for Contingent Valuation Validity Tests." Department of Agricultural, Resource, and Managerial Economics, Working Paper 96-16, Cornell University. Paper presented at the annual meetings of the American Agricultural Economics Association, Toronto.

Rondeau, D., W. D. Schulze, and G. L. Poe. 1998. "Voluntary Revelation of the Demand for Public Goods Using a Provision Point Mechanism." Journal of Public Economics forthcoming.

Rose, S., J. Clark, G. Poe, D. Rondeau, and W. Schulze. 1996. "Field and Laboratory Tests of a Provision Point Mechanism." Paper presented at the Economic Science Association Meeting, Tucson, AZ.

Rose, S., J. Clark, G. Poe, D. Rondeau, and W. Schulze. 1997. "The Private Provision of Public Goods: Tests of a Provision Point Mechanism for Funding Green Power Programs." Department of Agricultural, Resource, and Managerial Economics, Working Paper Series in Environmental and Resource Economics 97-02, Cornell University. Paper presented at the annual meetings of the Northeastern Agricultural and Resource Economics Association, June 1997, Sturbridge, MA.

Schulze, W.D. 1994. "Green Pricing: Solutions for the Potential Free Rider Problem." Paper prepared for Niagara Mohawk Power Corporation.

Schuman, H. 1996. "The Sensitivity of CV Outcomes to CV Survey Methods." In The Contingent Valuation of Environmental Resources, ed. D. Bjornstad and J.R. Kahn. Brookfield, VT: Edward Elgar.

Schuman, H., and S. Presser. 1981. Questions and Answers in Attitude Survey: Experiments on Question Form, Wording, and Context. San Diego: Academic Press.

Seip, K., and J. Strand. 1992. "Willingness to Pay for Environmental Goods in Norway: A Contingent Valuation Study with Real Payment." Environmental and Resource Economics 2:91-106.

Traugott, M., R. Groves, and J. Lepkowski. 1987. “Using Dual Frame Designs to Reduce Nonresponse in Telephone Surveys.” Public Opinion Quarterly 51:522-439.

Whittaker, D., J. Vaske, M. Donnelly, and D. DeRuiter. 1998. “Mail Versus Telephone Surveys: Potential Biases in Expenditure and Willingness-To-Pay Data.” Journal of Park and Recreation Administration forthcoming .

Wood, L.L., W.H. Desvousges, A.E. Kenyon, M.V. Bala, F.R. Johnson, R. Iachan, and E.E. Fries. 1994. “Evaluating the Market for ‘Green’ Products: WTP Results and Market Penetration Forecasts.” Center for Economics Research, Working Paper #4, Research Triangle Institute, NC.

Endnotes

1. Dillman (1991) discusses the three likely explanations for the presence of mode effects: social desirability bias, context effects, and “pace and control” effects stemming from differences in the answering process. We discuss only social desirability bias in this paper, as the expected effects of context and pace in this case are unclear, and in any case are thought to be less significant.

2. An alternative position, suggested by Michael Welsh of Christensen Associates, is that social desirability bias might act to lower hypothetical WTP and reduce hypothetical bias. In the presence of a researcher, respondents might take greater efforts to answer truthfully about how much they actually would be willing to pay if they perceive that such consideration and accuracy would please the researcher. In other words, they don’t want to be in a position where they might be perceived to be lying to the researcher.

3. From the perspective of CV validity testing the lack of price points other than $6/month is a limitation. A standard feature of dichotomous choice surveys is to use different price levels for different subsamples, allowing a WTP function or distribution to be estimated. Because the actual Green Choice program was only offered at $6/month, our comparisons are limited to this point estimate of WTP. As discussed in Balistreri *et al.* (1996), a single point estimate does not allow a separation of spread and location effects on mean WTP. Nevertheless, such comparisons do provide essential insights into the validity of CV [Cummings *et al.*, 1995].

4. In other circumstances, a mail survey might permit the use of visual aids or allow the researcher to convey more complex information than is practical over the phone. Since the same surveys were conducted over the phone and through the mail, with minor wording changes to make them amenable to each mode, specialized visual aids were not utilized.

5. For simplicity, except for WTP questions, we will restrict our comparisons to hypothetical phone and hypothetical mail treatments. Actual phone response patterns for non-WTP variables did not differ from hypothetical phone responses.

6. Although unconditional response rates do not differ, it is possible that conditional response patterns are not similar when demographic characteristics are accounted for. The hypothesis of equality for all coefficients was not rejected using likelihood ratio test comparisons of the logit response functions that included exogenous explanatory variables age, gender, education, and income (LR = 9.05 < 9.24 = $\chi^2_{1,\ 0.10}$) More specifically, a binary mail-phone variable was not found to be statistically different in the following logit estimation, with Pr(Yes) as the dependent variable:

$$\underset{(0.43)}{0.0079} - \underset{(0.20)}{0.12}*[\text{Phone}=1] - \underset{(0.007)}{0.022}*[\text{Age}] + \underset{(0.23)}{0.29}*[\text{College Grad}=1] + \underset{(3.9*10^{-6})}{9.2*10^{-6}}*[\text{Income}]$$

where numbers in () are asymptotic standard errors, n= 557, and only age and income were found to be significant at the 10% level or lower. Responses to “subjective” variables were not included in the list of explanatory variables because of their endogeneity to mode type (see Proposition 2).

7. See Poe *et al.* (1997) for further discussion.

Table 1. Response and Participation Rates by Sample

	Actual Phone Sign-Ups	Hypothetical Phone Sign Ups	Hypothetical Mail Sign Ups
Initial Sample Size	250	500	500
Removal from Sample Due to "List Errors"[a]	23	55	42
Removal from Sample Due to "Screening Questions"[b]	21	59	34
Adjusted Sample Size	206	386	424
Completed Surveys	145	275	285
Adjusted Response Rate	70.4%	71.2%	67.2%
Participation Rate	20.4%	30.6%	35.5%

a. "List Errors" are: undeliverables, incorrect or changed telephone number that could not be corrected by published telephone listings, and deceased.

b. “Screening Questions” excluded households who were not NMPC customers, those respondents not responsible for the electricity bill, and those respondents who recalled reading or hearing about the Green Choice program.

Table 2. Demographic Variables: Differences Between Phone and Mail Surveys

Variable	Chi-Squared[a] (n)	df	Phone Mean	Mail Mean
Age[b]	6.01 (539)	6	52.32	53.22
Gender	1.14 (549)	1	47.8% male	52.4% male
Income[c]	4.25 (499)	5	$35,747	$39,583
Occupation[d]	8.12 (540)	8	n.a.	n.a.
School[e]	10.97 (545)	7	13.7 Years	14.0 Years

a. *,**,*** Significant at 0.10, 0.05, and 0.01 levels, respectively.

b. Survey responses for age were continuous but converted to categories (21-30, 31-40, 41-50, 51-60, 61-70, 71-80, and above 80) for the Chi-Squared analyses.

c. Survey responses for household income were categorical as follows under $15,000, $15,000 to $30,000, $30,000 to $50,000, $50,000 to $75,000, $75,000 to $100,000, $100,000 to $150,000, $150,000 to $250,000, $250,000 or above. Because of small expected values (<5) the highest three categories were pooled for the Chi-Squared test. Mean response values were calculated from the midpoints of category values and no respondent selected the $250,000 or above category.

d. Two sets of categories (out of 11) were combined for occupation, one set because of small expected values (<5) and another (sales and services) because their substitutability had seemingly led to coding inconsistencies. The sales/service ratios were 6/19 for phone and 14/2 for mail, which contrasted markedly with the remaining eight categories. Since mail was self-coded and phone was coded using the first selected answer, inconsistencies were likely in similar categories like sales and service.

[e] Categorical response ranging from no school to post graduate.

Table 3. Subjective Variables: Differences Between Phone and Mail Surveys

Variable	Chi-Squared (n)[a]	df	Phone Mean	Mail Mean
Give To Environmental Causes	4.53** (546)	1	22.68%	15.52%
Rate Service[b] (1-10)	18.48*** (547)	6	8.45	7.98
Interest in Renewable Energy (1-10)	9.84 (519)	9	6.31	6.20
Interest in Planting Trees (1-10)[b]	17.97*** (525)	6	8.42	7.81

a. *,**,*** Significant at 0.10, 0.05, and 0.01 levels, respectively.

b. Response categories 1-4 were pooled because of low expected values (<5).

Table 4. Calibrated Hypothetical Sign-Ups with Comparison to Actual Sign-ups. [a,b]

	Actual Phone Sign-Ups (%)	Hypothetical Phone		Hypothetical Mail	
		Frequency	Estimated Sign-Up Rate (%)[c]	Frequency	Estimated Sign-Up Rate (%)[c]
Participation Rate	20.4	79	30.6**	97	35.5***
Certainty ≥ 5	--	75	29.1*	90	32.1**
Certainty ≥ 6	--	64	24.8	67	23.9
Certainty ≥ 7	--	54	20.9	60	21.4
Certainty ≥ 8	--	36	14.0*	43	15.4
Certainty ≥ 9	--	22	8.5***	24	8.6***
Certainty ≥ 10	- -	17	6.6***	17	6.1***

a. Sign-ups adjusted for certainty are counted as ‘yes’ if the respondent answered ‘yes’ to the dichotomous choice question at $6 per month and said that he/she would actually sign-up with a certainty level greater than or equal to X. Computations based on useable responses to both the dichotomous choice and the following calibration question: n_{Actual} = 142, $n_{Hypo. Phone}$ = 258, $n_{Hypo. Mail}$ = 273.

b. As noted in the text, none of the certainty adjusted pairs of hypothetical phone and hypothetical mail sign-up rates are significantly different at the 0.10 level.

c. *,**,*** Significant difference between the calibrated hypothetical and actual sign-ups at 0.10, 0.05, and 0.01 levels, respectively.

CHAPTER 6
Alternative Nonmarket Value-Elicitation Methods: Are the Underlying Preferences the Same?*

1. Introduction

Researchers who must resort to contingent choice survey methods to establish the social value of nonmarket environmental resources are by now familiar with the controversy over whether systematic biases in measured resource values can result from the investigator's choice of a value elicitation method.

All of the commonly used elicitation methods are intended to elicit the same underlying preferences. If data collected using different formats lead to statistically indistinguishable estimates of the parameters of the assumed preference function, then it is said that "procedural invariance" holds (Tversky, Sattath, and Slovic, 1988; Kahneman and Tversky, 1984) and therefore that the different elicitation methods are equivalent. The existing literature offers several pair-wise comparisons of elicitation methods, but the evidence is fragmented.

We employ a unique survey, designed specifically to allow simultaneous comparison of values elicited by six alternative methods that have been employed elsewhere in the literature. By ensuring that virtually everything about the survey instrument is strictly controlled across the different sub-samples involved and that the type of elicitation method is randomly assigned across respondents, we have an opportunity to directly compare the implied preference functions for respondents in each group. Unlike many past comparison studies, we can be confident that observed differences in values across methods do not stem instead from different descriptions of the environmental good, from distinctly different time frames for the survey, or from different populations being sampled.

Consider the situation where the researcher is willing to allow for differing error variances (heteroscedasticity across elicitation methods) in the stochastic portion of a very simple indirect utility function that is assumed to drive individuals' responses to valuation questions. In this case

***Authors: Trudy Ann Cameron, Gregory L. Poe, Robert G. Ethier, and William D. Schulze**

we show that it is not possible—for our data—to reject the hypothesis of identical underlying preference functions. This is an important result because if preferences are identical across value elicitation methods, then so will be expected willingness to pay for the good being valued.

In section 2, we situate this research in the context of prior literature on related issues. Section 3 describes our basic survey and its six variants, each one involving a different method for eliciting non-market values of an environmental good. In section 4, we introduce our simple indirect utility-difference model and its corresponding willingness to pay function, compatible in both their systematic and stochastic components, that allow us to specify completely conformable empirical models for the choice data from each elicitation method. The empirical results are presented in Section 5. We describe the results from using each elicitation method in isolation, then pair-wise, and then when all of the data are pooled in one encompassing model. Section 6 considers the implications of our work for subsequent research efforts in this vein, and section 7 concludes.

2. Review of Prior Findings

a.) Comparing Different Elicitation Methods

Many previous studies have undertaken pair-wise comparisons between different elicitation methods. These studies have contributed to our conventional wisdom about the differences that can typically be expected among the values of non-market goods derived using these different methods. For example, the prevailing finding in comparisons of dichotomous choice (DC) contingent valuation methods with open-ended (OE) methods suggests that the dichotomous choice methods produce estimates that tend to be larger--sometimes much larger. A recent example is Loomis et al. (1997), although Kramer and Mercer (1997) find little difference. Numerous comparisons are summarized in Schulze et al. (1996) and Brown et al. (1996). Some of the studies cited are Sellar et al. (1985), Boyle and Bishop (1988), Johnson et al. (1990), Walsh et al. (1992), Kealy and Turner (1993), and McFadden

(1994). The DC/OE willingness to pay ratio generally seems to range between 1.1 and 5, but with a number of exceptions. Huang and Smith (1998) review the differences between DC and OE methods using Monte Carlo methods and conclude that most of the difference seems to arise from specification errors common to the empirical models in the literature.

Other studies have compared dichotomous choice elicitation methods with payment card (PC) methods. These include Holmes and Kramer (1995), Ready et al. (1992) and Poe and Welsh (1996). The DC/PC willingness to pay ratio appears to range between 2.7 and 4.4 in these studies.

There are fewer instances of comparisons between multiple bounded (MB) elicitation methods and DC, PC, or OE methods. "Multiple bounded" is the term used here to describe an elicitation technique where the respondent is allowed to choose the *extent* to which he or she might be willing to pay. Sometimes, each respondent is also asked to rate their willingness to pay at each of several different "bid" values. Eliciting several choices from each respondent in this manner increases the available information about their preferences. Poe and Welsh (1998) have attempted some comparisons. The equivalence of MB and DC values is uncertain, but there is some evidence that MB values are moderately comparable to PC and OE value estimates.

Stated preference (SP) methods, now commonly associated with conjoint analysis methods originating in the marketing and transportation literatures (an referred to occasionally as “choice analysis” methods), ask respondents to choose among a set of scenarios that differ along several dimensions. Desvousges et al. (1983), Barrett et al. (1996), and Stevens et al. (1997) have explored these techniques in comparison with earlier CV formats and seem to find that SP value estimates exceed other types of CV estimates.

b.) Comparing Pairs of Elicitation Methods; Differing Variances

One significant antecedent for this paper is Boyle et al. (1996), which compares dichotomous choice with open-ended valuation responses, using independent samples and corresponding probit and tobit specifications allowing for differing error variances in the two sub-models. A common central tendency cannot be statistically rejected for two of their three pairs of data sets, but the estimated standard deviations are significantly different for all three pairs of data. Both the means and the standard deviations from the referendum-style samples exceed those from comparable open-ended data sets. These authors conclude that either open-ended questions underestimate values, or referendum-style questions overestimate them.

In a similar vein, Halvorsen and Soelensminde (1998) compare DC and OE responses collected from the same individuals and estimated simultaneously as a discrete/continuous bivariate probit-like specifications with a variety of alternative assumptions. They find heteroscedasticity across elicitation modes (and within the DC mode), but also different expected WTP. However, the fact that they do not employ split samples means that the endogeneity of the OE follow-up question may be contaminating the comparison of the two methods.

The present paper is differentiated from the Boyle et al. paper in that we employ the underlying utility-difference function as the basis from which *all* of our sub-model specifications are derived and we adopt the logistic error distribution assumption most common in random utility models. Furthermore, we combine not just pairs of samples, but six independent samples, all closely controlled except for their different elicitation methods.

c.) Comparing Stated and Revealed Preference Data with Different "Scales"

In the stated preference, or "conjoint analysis" literature, several papers have recognized the possibility of different scales of the latent variables underlying choices in different samples. Econometric tests of the equality of the scale parameters across samples have been conducted by Swait and Louviere (1993), by Adamowicz et al. (1994, 1997), Adamowicz et al. (1998) and by Boxall et al.

(1998). The most comprehensive citation is Hensher et al. (1999).

The idea of different scales across real and hypothetical elicitation methods also appears in a comment by Haab et al. (1998) concerning a paper by Cummings et al. (1997), wherein the incentive compatibility of real and hypothetical responses is assessed. Haab and his collaborators undertake to re-estimate the Cummings et al. empirical model allowing for different error dispersions. They accomplish the estimation and testing of models (with and without heteroscedasticity) using packaged maximum likelihood discrete choice models. This is accomplished by conducting a line-search across values of the unknown proportionality parameter for the error variances in the two samples.

Our paper differs from papers in the conjoint tradition in that our basic log-likelihood function differs across our independent samples in far more ways than just the magnitude of the error dispersion parameter. In the conjoint tradition, the type of elicitation methods usually differ only in that some choices are observed and some are hypothetical although completely analogous. We have one observed discrete choice, one analogous hypothetical discrete choice, but also four additional, very different hypothetical choices. All of these must be accommodated within one unified specification for rigorous statistical testing of the equivalence of the underlying preferences.

Our paper also differs from the Haab et al. strategy in that we avoid the device of a grid search and instead specify and estimate an appropriate log-likelihood function. With six pooled data sources, we normalize one dispersion parameter to unity and estimate dispersion parameters for the other data types as multiples of the first. These multiplicative dispersion factors are estimated jointly with the other model parameters in a full information maximum likelihood procedure.

3. Description of Survey Versions

Our data consist of responses to both telephone and mail surveys, all of which were conducted with the cooperation of the Niagara Mohawk Power Company (NMPC) in Erie County of New York

State. The topic of these surveys was the NMPC GreenChoice^{tm} program, wherein randomly selected households within NMPC's service territory were invited to consider either real or hypothetical additional charges on their utility bills in order to allow NMPC to plant trees and/or provide energy from renewable sources.[1] Within the telephone mode and within the mail mode, the survey instrument was identical except for the manner in which consumer values for the proposed GreenChoice program were elicited. Elicitation methods were assigned randomly across households within each survey mode. Thus, there can be a presumption that the only explanation for systematic differences in the implied preferences within a survey mode is these different elicitation methods.

The data for the mail survey were collected from a random sample of households with listed telephone numbers from the NMPC service territory within Erie county. An advance notification letter was sent to candidate households on November 14, 1996, and an initial copy of the mail survey with a $2 incentive payment was mailed November 18. A first follow-up survey mailing to non-respondents took place on December 5, and a second follow-up on January 23, 1997.

The telephone survey targeted the same population. Advance notification letters with a $2 incentive were used here as well. Computer assisted telephone interviewing commenced June 6, 1996, and ran through August 11. At least eight contact attempts were made for households in the intended sample who could not be reached initially. Telephone interviews averaged 13 minutes in length.[2]

Overall, adjusted response rates to the different survey variants ranged from about 64% to 69% in the mail surveys and were about 70% in the telephone survey. See the Appendix and Table A1 for a discussion of our non-response analyses and descriptive statistics on response proportions by survey variant.

---

[1] Ethier et al. (1977) use a portion of the data from this same study.

[2] The earlier time frame for the telephone interviews (summer 1996 versus late fall and winter) means that it is conceivable that preferences may have shifted seasonally. Thus there is a small chance that our actual choices from the telephone survey may not be as compatible with the mail surveys as the different versions of the mail survey are among

The survey instrument was titled "Clean Energy and You." The preamble was common to all version of the survey. First, the GreenChoice^{tm} program was introduced as a voluntary partnership between Niagara Mohawk and its residential customers, designed to reduce air pollution and improve the environment in "our local communities." The program involves two parts: (i) using renewable energy, and (ii) planting trees. First, the distinction between non-renewable and renewable energy sources was described. Potential renewable energy sources were identified as wind, solar power, gas recovered from landfill sites. It was pointed out that these energy sources do not produce air or water pollution, they will conserve resources, but they also tend to cost more than other types of power.

The second part of the GreenChoice^{tm} program, if implemented, would plant thousands of trees on public lands throughout upstate New York. These planting projects would be developed with American Forests, the nation's oldest citizen conservation group. Respondents were then informed about the role of trees as "natural air filters, absorbing carbon dioxide (a contributor to global warming) and releasing oxygen into the atmosphere. When planted near buildings, trees help conserve energy by providing shade in summer and windbreaks in winter."

Following the description of the GreenChoice^{tm} program, a key component of this exercise was the description of how the program would work. The "provision point mechanism" and "refund" plan were described as follows:

> The Green Choice program would be funded voluntarily. Customers who decided to join the program would pay an additional fixed fee each month on their Niagara Mohawk bill. This fee would not be tax-deductible. Customers could sign up or cancel at any time. While customers sign up, Niagara Mohawk would ask for bids on renewable energy projects.
>
> Enough customers would have to become Green Choice partners to pay for the program. For example, if $864,000 were invested in the first year, it would allow Niagara Mohawk to plant 50,000 trees and fund a landfill gas project. The gas project could replace all fossil fuel electricity in 1,200 homes. However, if after one year participation were insufficient to fund green Choice activities, Niagara Mohawk would cancel the program and refund all the money that was collected.

The survey variants for this study were designed specifically to allow comparisons across elicitation methods with minimum ambiguity. In the sections below, we provide details on the elicitation format for the portion of the questionnaire that differs across subsamples.

Survey Version 1: *(ACT) telephone actual purchase decision, $6 bid (Sample 1)*

After the common preamble, respondents were posed the following question:

> So far I've described the GreenChoice program, as well as the $6 per month it would add to your household's electrical bill, if you were to join.
>
> You may need a moment to consider the next couple of questions. Given your household's income and expenses, I'd like you to think about whether or not you would be interested in the GreenChoice program.
>
> If you decide to sign up, we will send your name to Niagara Mohawk, and get you enrolled in the program. All your other answers to this survey will remain confidential.
>
> Does your household want to sign up for the program at a cost of $6.00 per month?

Survey Version 2: *(MDC) dichotomous choice referendum, varying bids (Samples 2 through 8)*

For this survey version, there were seven samples, each one with a different bid value. For samples 2 through 8, the bids were $0.50, $1, $2, $4, $6, $9, and $12, respectively. For each of these survey variants (and for Survey Versions 3 through 6 discussed below as well), the common preamble to the survey was followed by the budget constraint reminder: "Given your household's income and other expenses, we would like you to think about whether or not you would be interested in joining the GreenChoice program." The MDC valuation question was posed as follows (e.g. in the case of the $4 bid).

> 10. Would your household sign up for the program if it cost you $4 per month?
> *(Please circle ONE response)*
>
> 1 Yes
> 2 No

Survey Version 3: *(OE) open-ended willingness to pay (Sample 9)*

For this survey version, respondents were asked to indicate the highest amount they would willingly pay for the good. The common preamble was followed by:

> 10. What is the highest amount, if anything, that your household would pay each month and still sign up for the program? *(Please fill in amount below)*
>
> **$ __________ per month**

Survey Version 4: (PC) payment card (Sample 10)

For this survey version, respondents were offered a payment card and were asked to circle the highest amount willingly paid, from which we infer that their true willingness to pay is greater than or equal to this amount, but strictly less than the next highest amount shown on the card. We thus use the amounts to form intervals and situate each response in a unique interval.[3]

> 10. What is the highest amount, if anything, that your household would pay each month and still sign up for the program?
>
> (Please circle the HIGHEST amount you would pay PER MONTH for the program.)

0	$.50	$1	$1.50	$2
$3	$4	$5	$6	$9
$12	$16	$20	$25	$35
$45	$55	$75	$95	more than $95

Survey Version 5: (MB) multiple bounded (Samples 11 through 13)

For this survey version, respondents were given five possible responses to each of 13 different bids and asked to circle the response that best characterized their degree of willingness to pay each bid value. These 13 bids differed for samples 11, 12 and 13. In the first sample (Sample 11), the middle

[3] There is some anecdotal evidence, from debriefing sessions with respondents who have faced payment cards, that respondents do not respond exactly to the instructions about the "highest amount" willingly paid. Instead, they may focus on the nearest amount. This interpretation is somewhat arbitrary, since we do not know who follows the instructions and who adopts their own strategy.

bid was $4. For this sample, the question format was as follows:

10. Would you joint the Green Choice program if it would cost you these amounts each month?

per month	Definitely No	Probably No	Not Sure	Probably Yes	Definitely Yes
10¢	A	B	C	D	E
50¢	A	B	C	D	E
$1	A	B	C	D	E
$1.50	A	B	C	D	E
$2	A	B	C	D	E
$3	A	B	C	D	E
$4	A	B	C	D	E
$6	A	B	C	D	E
$9	A	B	C	D	E
$12	A	B	C	D	E
$20	A	B	C	D	E
$45	A	B	C	D	E
$95	A	B	C	D	E

In the second multiple-bounded sample (Sample 12), the middle bid was halved, to $2.00. The full range of bids was $0.10, $0.25, $0.50, $0.75, $1, $1.50, $2, $3, $4, $6, $12, $35, $95. For the third multiple-bounded sample (Sample 13), the middle bid was tripled, to $12.00. The individual bids were $0.10, $0.50, $1, $2, $4, $8, $12, $16, $20, $25, $35, $55, $95. Note that care was taken to ensure that the lowest and highest bids were the same across Samples 11 through 13, so that respondents would not be cued differently by the *range* of bids on these three instruments, only by differences in the distribution of bids (if at all).

Survey Version 6: (SP) stated preference (Samples 14 and 15).

In this version of the survey, respondents are invited to make choices between pairs of possible programs (of different complexity and at different costs). The presence of other program alternatives may potentially confound a respondent's choice between Option C (paying for the standard scenario) versus Option A (not paying and not gaining this environmental enhancement). The five different program scenarios for this version are:

Option A: pay nothing, get no environmental goods
Option B: plant 50,000 trees
Option C: plant 50,000 trees, provide renewable energy to 1,200 homes
Option D: plant 100,000 trees, provide renewable energy to 1,200 homes
Option E: plant 100,000 trees, provide renewable energy to 2,400 homes

For the stated choice surveys, each of Options B through E is first compared pair-wise with Option A, which amounts to a set of binary discrete-choice referenda for each respondent. Next, these same respondents are invited to choose the most-preferred of the full set of five options (which still includes the do-nothing option).[4] The numbers of trees and houses involved for each of Options A through E for Samples 14 and 15 are identical across respondents, but the prices for each option differ between these two samples. For Sample 14, the prices are $0, $0.50, $2.00, $2.50, and $4.00. For Sample 15, the prices are $0, $2.00, $6.00, $8.00, and $12.00, as depicted below.

There are different costs associated with each of the proposed options. Increasing the number of trees planted or the number of homes serviced by renewable energy sources raised the cost of the Green Choice program. Costs per household per month are given below for each of the five options discussed on the previous pages. For each option the program would remain voluntary. Note that Option A, no Green Choice program, would cost $0 <u>per month</u>. It is not presented below.

11. Option B, planting 50,000 trees, would cost $2 each month. Would you be willing to pay $2 per month for Option B? *(Please circle ONE response)*

1 Yes
2 No
3 Don't Know

[4] Respondents were then invited to choose the next-most-preferred of these five options if their first choice was not available. However, respondents had difficulty with the idea that the "do nothing" option might not be available, so we elect not to assess whether these second-choices are consistent with the rest of the data.

12. Option C, planting 50,000 trees and providing renewable energy for 1,200 homes, would cost $6 per month. Would you be willing to pay $6 per month for Option C? (Please circle ONE response)

1 Yes
2 No
3 Don't Know

13. Option D, planting 100,000 trees and providing renewable energy for 1,200 homes, would cost $8 each month. Would you be willing to pay $8 per month for Option D?

1 Yes
2 No
3 Don't Know

14. Option E, planting 100,000 trees and providing renewable energy for 2,400 homes, would cost $12 each month. Would you be willing to pay $12 per month for Option E? *(Please circle ONE response)*

1 Yes
2 No
3 Don't Know

Given your household's income and other expenses, we would like you to think about whether or not you would be interested in joining the Green Choice program, and if so, which option you would prefer. Below are a number of ways the Green Choice program could be implemented, including the cost of the program each month. You have the opportunity to choose your most preferred option.

15. If the Green Choice program was made available, which option would be your first choice? *(Circle the LETTER of your FIRST choice on the list below)*

Option	Cost per Month	Number of Homes Fueled with Renewable Energy	Number of Trees Planted
A	$0	0	0
B	$2	0	50,000
C	$6	1,200	50,000
D	$8	1,200	100,000
E	$12	2,400	100,000

4. Theoretical and Econometric Modeling of Choices

Whereas many early comparisons of WTP values produced by different elicitation methods employed entirely separate models for separate data sets, our objective in this paper is to create one "grand unifying model" that subsumes all the different types of choice data produced by our different survey variants. Only when a common preference structure and stochastic specification underlie all of the choice models is it possible to combine them all in a single model. In what follows, we will describe the components of this unified model individually. Since we employed a split sample design, we will then be able simply to add up the components in a single log-likelihood function to be maximized with respect to a common set of utility parameters that show up (differently) in each component. In this way, utility parameters and error distribution parameters can be restricted or unrestricted across elicitation modes, and likelihood ratio test statistics can be used to conduct formal hypothesis tests regarding these utility and error distribution parameters.

a.) The Common Preference Model

There are potentially five distinct indirect utility function parameters in the simplest model we use to compare the preferences implied by respondents' choices under our six different elicitation methods. Each of these parameters will be assumed in this paper to be a simple constant, which implies a common preference function for all respondents.[5]

Recall that the full set of all five possible environmental enhancement scenarios are considered only in Version 6 of the survey, the stated preference (SP) variant. All of the other versions are concerned only with the choice between "doing nothing and paying nothing," (Option A) or choosing 50,000 trees and renewable energy for 1,200 houses at a price (Option C). Options B, D, and E

[5] If it is desired to allow preferences to vary systematically with demographic characteristics, for example, these constants can be generalized to functions of a set of explanatory variables (such as gender, age, etc.). We are pursuing this generalization in a subsequent paper.

(involving alternative numbers of trees and houses) appear only in Version 6.

The individual's level of indirect utility is presumed to depend upon the numbers of trees and houses affected, and the price of the proposed program. The four possible "do something" program alternatives can be captured by four dummy variables, B_i, C_i, D_i, and E_i, only one of which will be active for any program scenario.[6] Let V_i^1 be the level of indirect utility if one of these programs is chosen, and let V_i^0 be utility with option A (no program and no payment). Indirect utility with and without program participation can then be captured, in the simplest linear case, by:

$$(4.1) \qquad V_i^1 = \beta_0^* + \beta_1^* C_i + \beta_2^* B_i + \beta_3^* D_i + \beta_4^* E_i + \beta_5^* (Y_i - price_i) + u_i^1$$

$$V_i^0 = \beta_0^* + \beta_5^* (Y_i) + u_i^0$$

Consistent with the usual assumptions of random utility choice models, we will assume that the error term, u_i^1 or u_i^0, has an extreme value distribution.

We will also assume that β_1^* through β_4^* are strictly positive, implying that the commodities being valued are "goods," not "bads."[7] We will also assume that β_5^* is positive, which is required for rationality, in the sense that indirect utility should not *decrease* with income (or *increase* with price).

We restrict β_5^* to be strictly positive by estimating β_5^* as $\exp(\beta_5)$, where β_5 is a (potentially systematically varying) parameter which can take on any value dictated by the data. This restriction is empirically necessary primarily to ensure the theoretical plausibility of the models corresponding to Versions 3 and 4 of the survey (open-ended willingness to pay, and payment card models). We can similarly restrict the other ("intercept") parameters of the indirect utility function to be positive by

---

[6] We capture the four different do something program alternatives as dummy variables because there are only two distinct levels for the numbers of trees and the number of houses. This is insufficient information to allow utility to vary continuously with the numbers of trees and houses.

[7] In all but Survey Version 6, only Option C is being compared to Option A, so we are estimating only β_1^* and β_5^* (or, in more general specifications, systematically varying versions of these parameters). The dummy variables $B_i = D_i = E_i = 0$ for all observations for these versions.

estimating each β_j^* as $\exp(\beta_j)$, j = 1,..,4. Then the underlying parameter β_j can take on any value.[8]

An individual's choice regarding program participation is assumed to depend on whether the indirect utility difference (between participation and non-participation) is positive. At least three different alternatives can be considered for parameterization of the indirect utility-difference function that drives individual choices: scalar unconstrained parameters, scalar parameters constrained to be positive, and systematic varying parameters constrained to be positive. We illustrate for an individual's choice between Option C and the do-nothing alternative Option A:

(4.2a) $$(V^1 - V^0)_i = \beta_1^* - \beta_5^* \text{ price}_i + e_i \text{ , or}$$

(4.2b) $$(V^1 - V^0)_i = \exp(\beta_1) - \exp(\beta_5) \text{ price}_i + e_i \text{ , or}$$

(4.2c) $$(V^1 - V^0)_i = \exp(x_{1i}'\beta_1) - \exp(x_{5i}'\beta_5) \text{ price}_i + e_i \text{ ,}$$

where $e_i = u_i^1 - u_i^0$ is distributed logistic$(0,\kappa)$. In this paper, we focus upon variant (4.2b), the case of scalar parameters with sign restrictions.[9] Note that variant (4.2b) is simply variant (4.2c) with x_{1i} and x_{5i} equal to a constant term.

Vital to the pooled-data model in this paper is the correspondence between the indirect utility-difference function (which drives the discrete choices) and the continuous maximum willingness-to-pay (WTP) function (which forms the basis of the open-ended and payment card responses). If we take the version of the utility difference in (4.2b), set it equal to zero, and solve for price_i, this dollar value will represent the predicted maximum willingness to pay by any individual individual in the sample. The formula for fitted willingness-to-pay for Option C can therefore be expressed quite simply as:

---

[8] Again, this is particularly important if β_j is to be converted to a systematic varying parameter that depends on the observed levels of respondent attributes. In counterfactual simulations, we do not wish to find predicted values of β_j* that are less than zero. While the indirect utility gleaned from the various combinations of the two environmental goods (trees and houses) may be very small (or even essentially zero), we choose to preclude the possibility that it might be negative.

[9] The sign restrictions are not binding, but this convention facilitates comparison of the present results with those in a subsequent paper which employs systematically varying parameters.

(4.3) $$WTP_i = price_i = \exp(\beta_1)/\exp(\beta_5) + e_i/\exp(\beta_5).$$

The first term in this expression is constrained to be positive, but the transformed error term is a scaled version of the underlying logistic $(0,\kappa)$ error, for which the support is the entire real line.[10]

It is possible to interpret the fitted conditional distribution of WTP_i as the distribution of individual WTP values in the portion of the population represented by this observation. Thus, the fact that the error term is unbounded may influence our strategies for calculating predicted expected WTP. That portion of the fitted conditional density in the negative domain could be changed to a point mass at zero before the expected value is calculated. (This is akin to the interpretation of expected values in standard Tobit models.)

b.) Survey Version 1: (ACT) actual dichotomous choice referendum, \$6 bid (Sample 1)

Each individual in this sample is presented with the same \$6 bid value and invited actually to pay this amount for the scenario described (Option C, the main option). Indirect utility if this option is chosen is given by V_i^1; if it is not selected, indirect utility is V_i^0:

(4.5) $$V_i^1 = \beta_0^* + \exp(\beta_1)(1)+\exp(\beta_2)(0)+\exp(\beta_3)(0)+\exp(\beta_4)(0)+\exp(\beta_5)\,(Y_i - price_i) + u_i^1$$

$$V_i^0 = \beta_0^* + \exp(\beta_5)(Y_i) + u_i^0$$

Since Options B, D, and E are not being considered in this sample, the indirect utility-*difference*, $V_i^1 - V_i^0$, which drives this pair-wise choice can be simplified considerably:

(4.6) $$(V^1 - V^0)_i = \exp(\beta_1) - \exp(\beta_5)\, price_i + e_i = Z_{1i} + e_i$$

where $e_i = (u_i^1 - u_i^0)$, and the coefficients β_2 through β_4 do not appear. Since $price_i$ does not vary across

[10] If x_{5i} in specification variant (4.2c) contains any more than just a constant term, this error term will furthermore be heteroscedastic across individuals.

this sample (it is \$6 for everyone), $\exp(\beta_1)$ and $\exp(\beta_5)$ cannot be separately identified without additional information (such as that gleaned from Survey Version 2 or other versions). If x_1 and x_5 each consist solely of an "intercept" term, all that can be estimated is the sample average value of the index $Z_{1i} = \exp(\beta_1) + 6\exp(\beta_5) = b_1$. As usual for binary discrete choice models employed on a single sample, the scale of measurement of this indirect utility difference must be standardized by the unobserved logistic error dispersion parameter, κ_1, which is equivalent to normalizing κ_1 to unity.

If a respondent indicates they are willing to pay the bid amount, then let $I_{1i}=1$, otherwise $I_{1i}=0$. The contribution of Survey Version 1 to the log-likelihood function is therefore:

$$(4.7) \qquad \text{Log } L_1 = \sum_{i=1}^{n_1} I_{1i} \log\{\exp(Z_{1i}/\kappa_1)/[1+\exp(Z_{1i}/\kappa_1)]\} + (1-I_{1i}) \log\{1/[1+\exp(Z_{1i}/\kappa_1)]\}$$

Note that if $\kappa_1 = 1$, then Log L_1 is just the familiar binary logit discrete choice model.[11]

c.) Survey Version 2: (MDC) hypothetical dichotomous choice, varying bids (Samples 2 through 8)

Each individual in this sample is presented with a <u>different</u> bid value and likewise invited to indicate whether they would be willing to pay this amount for Option C. The indirect utility with and without Option C in this case is identical to that in equations (4.5). The important difference is that price_i now varies across observations. The indirect utility-difference index Z_{1i} above is thus replaced by $Z_{2i} = \exp(\beta_1) - \exp(\beta_5)\,\text{price}_i$. Thus, the indirect utility-difference that drives this pairwise choice will be:

$$(4.8) \qquad (V^1 - V^0)_i = Z_{2i} + e_{2i}$$

---

[11] If $\beta_1{}^*$ and $\beta_5{}^*$ are constants, a single parameter $Z_{1i} = b_1$ can readily be estimated. However, if we are assuming systematically varying preferences with sign constraints such that $\beta_1{}^* = \exp(x_i'\beta_1)$ and $\beta_5{}^* = -\exp(x_i'\beta_5)$, it does not seem possible to collapse the parameter space in any simple way. Of course, if we render $\beta_1{}^* + 6\beta_5{}^*$ into a *single* systematically varying index by substituting $x_i'\gamma$, instead of using *two* indexes in the form $\exp(x_i'\beta_1)$ $6\exp(x_i'\beta_5)$, it is possible to determine whether the probability that a respondent is willing to pay the \$6 bid differs systematically across individuals with different characteristics. This model, however, is not nested within any of the other models considered in our broader study.

where u_{2i} is distributed logistic$(0,\kappa_2)$, with κ_2 not necessarily identical to κ_1. The coefficients $\exp(\beta_2)$ through $\exp(\beta_4)$ still do not appear. Now, however, since $price_i$ *does* vary in this sample, $\exp(\beta_1)$ and $\exp(\beta_5)$ can be separately identified.

With $I_{2i} = 1$defined analogously to I_{1i}, indicating that the respondent is willing to pay the offered price for the basic scenario, and $I_{2i} = 0$ if the respondent is unwilling to pay this amount, the contribution to the log-likelihood function for Survey Version 2 is:

$$\text{(4.9)} \qquad \text{Log } L_2 = \sum_{i=1}^{n_2} \quad I_{2i} \log\{\exp(Z_{2i}/\kappa_2)/[1+\exp(Z_{2i}/\kappa_2)]\} + (1-I_{2i}) \log\{1/[1+\exp(Z_{2i}/\kappa_2)]\}$$

The log-likelihood function Log L_2 is again the familiar binary logit discrete choice log-likelihood, except for the κ_2 parameter which allows the dispersion to differ from that for other survey versions, unless Survey Version 2 data are the only data being used, in which case we normalize κ_2 to unity.

d.) Survey Version 3: open-ended willingness to pay (Sample 9)

This version of the survey invites respondents to directly state their willingness to pay for the offered scenario. If behavior is consistent with simple economic theory of utility maximization, and if our model is appropriate, this willingness to pay should be that dollar price which would make the respondent indifferent between paying the price and getting the offered scenario, or not paying and not getting the scenario.

The relevant indirect utility difference should be rendered zero by the open-ended willingness to pay amount, WTP_i. Thus we should have:

$$\text{(4.10)} \qquad (V^1 - V^0)_i = \exp(\beta_1) - \exp(\beta_5)\, price_i + e_{3i}$$

Solving for $WTP_i = price_i$ such that $(V^1 - V^0)_i = 0$ yields the expression given in equation (4.3) above.

(4.11) $$WTP_i = \exp(\beta_1)/\exp(\beta_5) + e_{3i}/\exp(\beta_5).$$

When this sample is combined with other samples, it may again be prudent to allow the error term e_{3i} to be distributed logistic$(0,\kappa_3)$, where κ_3 can differ from κ_1 and κ_2.

For the open-ended elicitation method, respondents presumably provide a value for WTP_i. The conditional expected value of this distribution is $\exp(\beta_1)/\exp(\beta_5)$ and the dispersion parameter is $\kappa_3/\exp(\beta_5)$. Define Z_{3i} fundamentally differently from Z_{1i} or Z_{2i}, by using it now to denote the "standardized" value of WTP_i:

(4.12) $$Z_{3i} = [\ WTP_i - \exp(\beta_1)/\exp(\beta_5)\ ] \ /\ [\ \kappa_3/\exp(\beta_5)\ ]$$

To simplify the notation in what follows, define $s_3 = \kappa_3/\exp(\beta_5)$. (If β_5 were to be generalized to $x_{5i}'\beta_5$, then s_3 would become s_{3i}.)

The assumption of logistically distributed regression errors can be adapted to a Tobit-like regression-by-maximum-likelihood context. The reason a Tobit-like model is indicated is because there is likely to be some heaping of reported WTP_i at zero, since negative values are not intuitively acceptable. Define $POS_i = 1$ if a strictly positive value of WTP_i is reported for observation i. If a zero WTP_i is reported, then $POS_i = 0$. The contribution of the responses to Survey Version 3 to the log-likelihood function is:

(4.13) $$\text{Log } L_3 = \sum_{i=1}^{n_3} \quad POS_i\ \{Z_{3i} - \log(s_3) - 2*\log[1 + \exp(Z_{3i})]\}$$
$$[1 - POS_i] \log\{\exp[Z_{3i}/(1+ \exp(Z_{3i}))]\}$$

This component of the overall log-likelihood function can be viewed as an analog to the familiar Tobit log-likelihood, but based on the logistic error distribution rather than the more common normal distribution.

e.) Survey Version 4: payment card (Sample 10)

This version of the survey generates interval data for the true but unobserved WTP_i value. As in the open-ended case, the latent variable we must model is WTP_i. It can be defined as in equation (4.7), but we will now substitute e_{4i}, distributed logistic$(0,\kappa_4)$, where $\kappa_4 \neq \kappa_j$, $j \neq 4$. Analogous to the normal-error payment card model used by Cameron and Huppert (1989), let t_{ui} be the upper bound of the payment card interval chosen by the respondent, and let t_{li} be the associated lower bound. Then define:

(4.14) $$Z_{ui} = [t_{ui} - \exp(\beta_1)/\exp(\beta_5)\ ] \,/\, [\ \kappa_4/\exp(\beta_5)\ ], \text{ and}$$

$$Z_{li} = [t_{li} - \exp(\beta_1)/\exp(\beta_5)\ ] \,/\, [\ \kappa_4/\exp(\beta_5)\ ].$$

If we treat the true but unknown value of WTP_i as the conditional expected value of WTP_i, the log-likelihood function involves the difference in the cumulative densities between the standardized upper bound (Z_{ui}) and the standardized lower bound (Z_{li}). For the assumed underlying logistic density function, the cumulative densities are $P_{ui} = 1/[1+\exp(-(Z_{ui}))]$ and $P_{li} = 1/[1+\exp(-(Z_{li}))]$. For our model with homogeneous preferences, the contribution to the log-likelihood of the observations from Survey Version 4 is:

(4.15) $$\text{Log } L_4 = \sum_{i=1}^{n_4} \log\{P_{ui} - P_{li}\}.$$

Note that if we are dealing with the lowest interval, we will use just $\log\{P_{ui}\}$; with the highest interval, we will use just $\log\{1-P_{li}\}$.

This component of the overall log-likelihood can thus be characterized as a variant of the common interval-data model, adapted to an underlying logistic error distribution instead of the usual normal distribution.

f.) Survey Version 5: multiple bounded (Samples 11 through 13)

Respondents receiving Survey Version 5 were presented with 13 different bid values and asked to indicate (in categories) how likely they would be to be willing to pay each of these amounts. The intuitive framework for analyzing these responses is a multi-category generalization of the binary discrete choice referendum that applies for Survey Versions 1 and 2—i.e. an ordered logit framework instead of a binary logit framework.

As for the first two survey versions, the indirect utility difference associated with paying the price and obtaining the scenario described (versus not paying and not obtaining the scenario) is presumed to drive the ordered categorical response to each question on this survey. Again, let the relevant indirect utility difference be:

$$(4.16) \qquad (V^1 - V^0)_i = \exp(\beta_1) - \exp(\beta_5)\, price_i + e_{5i} = Z_{5i} + e_{5i}$$

Note that we again allow the conditional error variance again to differ from that for other versions of the survey if other samples are being used in conjunction with this one.

Respondents to this survey variant chose one of five levels of likelihood that they would be willing to pay each of the 13 different bid amounts. Let Y_i=1 if the respondent chooses the "Definitely Yes" (yes) response, zero otherwise. Let H_i=1 if the respondent chooses the "Probably Yes" (high) response, zero otherwise. The indicators M_i, L_i, and N_i are defined similarly for the "Not Sure" (medium), "Probably No" (low) and "Definitely No" (no) responses.

For an ordered logit model with five response categories, there are four standardized estimated thresholds. Generally, researchers arbitrarily set one interval threshold to zero (usually the lowest one) since the location and scale of the underlying "propensity to be willing to pay" variable are unknown. But these data will sometimes be used in conjunction with other samples from this study, (in particular, the binary logit data from Survey Versions 1 and 2). In this case, we would expect the "zero" level of the latent propensity variable in this ordered logit to lie somewhere in the middle interval (i.e. in the "Not Sure" category).

Let the thresholds between the five intervals (no, low, medium, high and yes) be denoted as α_0, α_1, α_2, and α_3, respectively. Like the slope parameters in the binary choice models, these thresholds are known only up to a scale factor consisting of the dispersion parameter of the error term in this particular submodel, namely κ_5. For *multiple* samples, then, we are able to estimate α_0/κ_5, α_1/κ_5, α_2/κ_5, and α_3/κ_5, and κ_5 can itself be estimated separately as a multiple of some κ_j normalized to unity for some other sample j. For Survey Version 5 used alone, we must normalize $\alpha_0/\kappa_5 = 0$ and $\kappa_5 = 1$.

In the most general case, then, we can define five probabilities, one associated with each of the five categories within which each individual may have responded (for each of the thirteen choice opportunities afforded respondents to this survey variant):

$$
\begin{aligned}
(4.17)\qquad & PY_i = 1/[1+\exp(\alpha_3/\kappa_5 - Z_{5i})], \\
& PH_i = \exp(\alpha_3/\kappa_5 - Z_{5i})/[1+\exp(\alpha_3/\kappa_5 - Z_{5i})] - \exp(\alpha_2/\kappa_5 - Z_{5i})/[1+\exp(\alpha_2/\kappa_5 - Z_{5i})], \\
& PM_i = \exp(\alpha_2/\kappa_5 - Z_{5i})/[1+\exp(\alpha_2/\kappa_5 - Z_{5i})] - \exp(\alpha_1/\kappa_5 - Z_{5i})/[1+\exp(\alpha_1/\kappa_5 - Z_{5i})], \\
& PL_i = \exp(\alpha_1/\kappa_5 - Z_{5i})/[1+\exp(\alpha_1/\kappa_5 - Z_{5i})] - \exp(Z_{5i})/[1+\exp(Z_{5i})], \text{ and} \\
& PN_i = \exp(\alpha_0/\kappa_5 - Z_{5i})/[1+\exp(\alpha_0/\kappa_5 - Z_{5i})].
\end{aligned}
$$

The contribution to the log-likelihood function of the set of m=1,…,13 categorical responses for complete observations[12] from Survey Version 5 can then be expressed as:

[12] However many completed responses are provided to these thirteen questions by respondents receiving Survey Version 5, we scale the total contribution to the log-likelihood so that each respondent has a weight equal to respondents receiving other versions of the survey.

(4.18)

$$\text{Log } L_5 = \sum_{i=1}^{n_5} (1/13) \sum_{m=1}^{13}{}_1 \{ Y_i \log[PY_i] + H_i \log[PH_i] + M_i \log[PM_i] + L_i \log[PL_i] + N_i \log[PN_i] \}$$

In this paper, we constrain the vector of α parameters to be identical across all thirteen questions (or portion thereof answered by any respondent). We also estimate just a single κ_5 parameter for all thirteen questions. Generalizations, including exploitation of the panel nature of the data in this subscample, are being pursued in a subsequent paper.

g.) Survey Version 6: stated preference (Samples 14 and 15).

This version of the survey introduces, as distractors, three other possible scenarios in addition to the single binary choice scenario that is offered to respondents in Versions 1 through 5 of the survey. Of interest again is the indirect utility derived from each of the available scenarios. There are two samples for this version. One subsample was posed a lower set of prices for the set of scenarios, and the other was posed a higher set of prices. We will discuss the low-price instrument first (sample 14).

Rather than just V^1 and V^0, as before, we will now have five different indirect utility levels. Let these be denoted V^a (called V^0 above), V^b, V^c (called V^1 above), V^d, and V^e. Indirect utility under each of these choices can then be expressed in general as:

(4.19) $$V^j_i = \beta_0^* + \beta_1^* C^j_i + \beta_2^* B^j_i + \beta_3^* D^j_i + \beta_4^* E^j_i + \beta_5^* (Y_i - \text{price}^j_i) + u_i^j$$

and in particular as:

(4.20)
$$V^A_i = \beta_0^* + \beta_1^* (0) + \beta_2^* (0) + \beta_3^* (0) + \beta_4^* (0) + \beta_5^* (Y_i - \$0.00) + u^A_i = Z_{Ai} + u^A_i$$
$$V^B_i = \beta_0^* + \beta_1^* (0) + \beta_2^* (1) + \beta_3^* (0) + \beta_4^* (0) + \beta_5^* (Y_i - \$0.50) + u^B_i = Z_{Bi} + u^B_i$$
$$V^C_i = \beta_0^* + \beta_1^* (1) + \beta_2^* (0) + \beta_3^* (0) + \beta_4^* (0) + \beta_5^* (Y_i - \$2.00) + u^C_i = Z_{Ci} + u^C_i$$
$$V^D_i = \beta_0^* + \beta_1^* (0) + \beta_2^* (0) + \beta_3^* (1) + \beta_4^* (0) + \beta_5^* (Y_i - \$2.50) + u^D_i = Z_{Di} + u^D_i$$
$$V^E_i = \beta_0^* + \beta_1^* (0) + \beta_2^* (0) + \beta_3^* (0) + \beta_4^* (1) + \beta_5^* (Y_i - \$4.00) + u^E_i = Z_{Ei} + u^E_i.$$

Since sample 15 has higher prices ($0.00, $2.00, $6.00, $8.00, and $12.00), there is independent variation across the pooled samples 14 and 15 in the prices faced by respondents.

We will again allow for a different (common) error dispersion parameter from those that apply to the other samples. It would be possible to specify different dispersion parameters for each of the four pairwise choices made by respondents between each of Options B through E and Option A. Due to a relatively small numbers of observations, however, we elect to impose the restrictions $\kappa_{BA} = \kappa_{CA} = \kappa_{DA} = \kappa_{EA}$. We will refer to this common dispersion parameter as κ_{6P}, where the P denotes "pairwise choices."

For this survey version, however, respondents were also asked to choose their most-preferred option among Options A through E. This is a five-alternative multiple choice model that we will model as a multinomial logit choice. The error dispersion for this choice context could be different again, so we will allow for a separate parameter κ_{6J}, with the J subscript denoting the "joint choice."

For each of the pairwise choices of Options B through E against simply the do-nothing Option A, the probabilities will be:

(4.21)
$$
\begin{aligned}
P_{Bi} &= \exp(Z_{Bi}/\kappa_{6P}) / [\exp(Z_{Ai}/\kappa_{6P}) + \exp(Z_{Bi}/\kappa_{6P})] \\
P_{Ci} &= \exp(Z_{Ci}/\kappa_{6P}) / [\exp(Z_{Ai}/\kappa_{6P}) + \exp(Z_{Ci}/\kappa_{6P})] \\
P_{Di} &= \exp(Z_{Di}/\kappa_{6P}) / [\exp(Z_{Ai}/\kappa_{6P}) + \exp(Z_{Di}/\kappa_{6P})] \\
P_{Ei} &= \exp(Z_{Ei}/\kappa_{6P}) / [\exp(Z_{Ai}/\kappa_{6P}) + \exp(Z_{Ei}/\kappa_{6P})]
\end{aligned}
$$

We need indicators for a respondent's choice in each of these four pairwise comparisons. Let $I_{Bi}=1$ if scenario B is chosen over A by respondent i, $I_{Bi} = 0$ otherwise. Similarly for I_{Ci}, I_{Di} and I_{Ei}. For each alternative j = B, C, D, and E, the contribution to the log-likelihood function will be

(4.22)
$$
\text{Log } L_{6P} = \sum_{i=1}^{n_{6P}} \Sigma_{j=B,C,D,E} \; \{ \; I_{ji} \log(P_{ji}) + (1 - I_{ji}) \log(1 - P_{ji}) \; \}
$$

The second type of choice question posed to respondents who received the stated preference (SP) survey instrument asked them to choose their most-preferred option from the set of options A through E. We will allow for a separate dispersion parameter for this choice as well, denoted κ_{6J}. Simplify the notation to follow by defining $SUMP_i$ as $\exp(Z_{Ai}/\kappa_{6J}) + \exp(Z_{Bi}/\kappa_{6J}) + \exp(Z_{Ci}/\kappa_{6J}) + \exp(Z_{Di}/\kappa_{6J}) + \exp(Z_{Ei}/\kappa_{6J})$. Then under the logistic model, the probabilities of choosing each specific alternative from this set of five are given by:

$$(4.23)\qquad \begin{aligned} P_{A}'_i &= \exp(Z_{Ai}/\kappa_{6J})/SUMP_i \\ P_{B}'_i &= \exp(Z_{Bi}/\kappa_{6J})/SUMP_i \\ P_{C}'_i &= \exp(Z_{Ci}/\kappa_{6J})/SUMP_i \\ P_{D}'_i &= \exp(Z_{Di}/\kappa_{6J})/SUMP_i \\ P_{E}'_i &= \exp(Z_{Ei}/\kappa_{6J})/SUMP_i \end{aligned}$$

As indicator variables for this top choice among the five possibilities, define $I_{Ai}' = 1$ if scenario (A) is most-preferred, $I_{Ai}' = 0$ otherwise. Similarly, define I_{Bi}', I_{Ci}', I_{Di}', and I_{Ei}'. The contribution to the log-likelihood of this first-choice program from among the five alternatives is then:

$$(4.24)\qquad \text{Log } L_{6J} = \sum_{i=1}^{n_{6J}} I_{Ai}'\log(P_{Ai}') + I_{Bi}'\log(P_{Bi}') + I_{Ci}'\log(P_{Ci}') + I_{Di}'\log(P_{Di}') + I_{Ei}'\log(P_{Ei}')$$

It again seems appropriate that a single respondent to Survey Version 6 should have, in total, only unit weight in determining the maximized value of the log-likelihood function using pooled data. Thus, the number of pieces of choice information extracted from each respondent is used to divide the total contribution of each respondent to the overall log-likelihood function. For complete responses, for example, we would use $\text{Log } L_6 = (\text{Log } L_{6P} + \text{Log } L_{6J})/6$.

h. The complete specification

Due to the independence of the six samples, the log-likelihood function for the pooled data is just the sum of the six component likelihood terms. The parameters are found by maximizing:

(4.25) $$\text{Log } L = \text{Log } L_1 + \text{Log } L_2 + \text{Log } L_3 + \text{Log } L_4 + \text{Log } L_5 + \text{Log } L_6$$

The same utility parameters appear in each of the six components, and can be restricted or left unrestricted across the six log-likelihood terms as desired. The κ_j dispersion factors can be restricted to unity (making all dispersions the same as the (unidentified) normalized error dispersion for the numeraire sample. Alternately, all but one of the κ_j parameters representing dispersion factors for each survey version can be freely estimated.

5. Empirical Results

In this paper, we maintain the hypothesis of homogeneous preferences, in the sense that preference function parameters do not differ systematically across individuals according to their sociodemographic characteristics. Provided that dispersions are allowed to differ by elicitation method, we are unable to reject the restriction of common preference parameters across any pair of elicitation methods or even across all six methods. In demonstrating this result, we begin by considering the results from independent estimation, then pairwise estimation. We then consider a fully restricted model, and finally examine a less-restrictive model with preference parameters constrained but dispersion parameters allowed to differ.

a.) Independent Estimates based on Different Methods with Separate Samples

First, we can consider the consequences of using each of our individual samples, with its own elicitation method, in an entirely separate empirical model. The first column of Table 1 provides these estimates. Of course, the 1-ACT sample, by itself, cannot produce a point estimate of WTP because the bid value does not vary across respondents. But the fact that the three implicit parameters in this model cannot be separately identified does not prevent us from maximizing the log-likelihood associated with this specification and using this unconstrained model for comparison with the achievable log-likelihood values for other pooled-data models. Note that all other elicitation methods do allow us to produce a point estimate of the expected willingness to pay produced by that elicitation method.

Recall that our theoretical and statistical specifications for the indirect utility-difference (and thus the corresponding WTP function) for each sub-sample are entirely conformable, unlike the case for the underlying specifications in many earlier comparisons across elicitation methods. The functional form for the systematic portion of each model is consistent with the same underlying utility function, and the stochastic structures assumed for maximum likelihood estimation also correspond exactly across data types.[13]

What we see in the first column of Table 1 is familiar to researchers who have compared alternative elicitation methods: the point estimates of WTP differ rather markedly across our different samples--an outcome (in this case) attributable exclusively to the different elicitation methods that were used, since all other features of the sampling frame, survey instrument, and survey timing were controlled as rigorously as possible.

b.) Pairwise Comparisons across Individual Value-Elicitation Methods

Most previous studies which have drawn comparisons between alternative elicitation methods have considered different methods two at a time. Table 1 also reports selected results for all possible *pairs* of samples. In each comparison cell there are two sub-cells. First, we report key results when *all* parameters are constrained to be identical across the two samples. Then we report results when the utility parameters are constrained but the error dispersions are allowed to differ. The reported results

[13] It is possible, of course, that entirely ad hoc functional forms for the systematic and stochastic portions of these models might yield more-similar point estimates of WTP across methods. Then, however, there would be no rigorous method for conducting statistical tests of the equivalence of preferences.

are (i) the maximized value of the log-likelihood function for the joint model, (ii) the fitted point estimate of WTP for Option C based on the joint model, (iii) the number of estimated parameters in the joint model, and (iv) the aggregate number of (unweighted) choice occasions used to provide these estimates. (Sometimes more than one choice per individual is observed and used, although the total weight for each individual remains unitary). The final item reported is the factor of proportionality for the dispersion parameter κ_j for the non-numeraire sample in each pair. If the log of this value (the statistical quantity actually estimated) is statistically different from zero at the 5% level, there is an asterisk associated with this point estimate. Note that if $\log(\kappa)=0$, then $\kappa=1$, so that there is no difference between the two samples in the scale of the dispersion parameter.

The most important question with respect to each pair-wise comparison concerns whether the utility parameters are statistically the same across elicitation methods. If the sum of the log-likelihood values for each sub-sample (estimated separately) exceeds by a sufficient amount the maximized log-likelihood for a joint model with all corresponding parameters constrained to be identical, then we can reject the parameter restrictions in the constrained model. However, we are watching for evidence that freeing up the error variances means that the assumption of common corresponding utility parameters cannot be rejected. When the maximized log-likelihood appears in **bold face**, the restrictions in the joint model *are* rejected at the 5% level; when the log-likelihood appears in *italics*, the restrictions of the joint model *cannot* be rejected at the 5% level. Table 1 reveals that there is no pair-wise comparison for which *identical utility parameters* can be rejected, providing the error dispersions are allowed to differ across the two samples.

Why are we interested in the consequences of pooled data with parameter restrictions? If the indirect utility function parameters are in fact *the same* across samples with different elicitation methods, then the implied expected willingness to pay, $\exp(\beta_1)/\exp(\beta_5)$, is also identical across these samples. If we were to rely on separate-sample estimates of the utility parameters, and were to compute fitted willingness to pay for each of these different samples, we would get the apparently conflicting implications captured by the WTP estimates in the first column of Table 1. For Survey Version 1, where the threshold does not vary across respondents, it is not possible to recover a point estimate of WTP. However, for the remaining 5 samples, the estimates vary rather widely: \$2.84, \$1.30, \$2.09, \$3.08, and \$3.45 for samples 2 through 6 respectively. These are the types of discrepancies that have led to criticism of the robustness of value estimates across alternative hypothetical valuation methods.

c.) Pooled Model: All Samples

Table 2 reports two sets of results for cases where the data from all six survey versions are pooled. First, we estimate a common set of parameters for the indirect utility function that might be presumed to underlie all of these choices.

As in the pairwise comparisons considered above, we wish to compare the indirect utility-difference function implied by each sample (used individually) with the indirect utility-difference function estimated when preferences are constrained to be identical across all samples. Under an assumption of homogeneous preferences across individuals within a sample, the sum of the separate log-likelihood functions for the independent samples reported in the first column of Table 1 is -2712.768. The corresponding joint model, where all utility and dispersion parameters are restricted to be the same, achieved a maximized log-likelihood of only -2784.699, which clearly rejects these combined restrictions. However, if the utility parameters *remain* restricted, but we allow the dispersion parameters to differ across samples, the log-likelihood climbs all the way back to -2714.733, despite the large number of cross-sample indirect utility parameter restrictions remaining. These utility parameter restrictions therefore cannot be rejected. This finding is the key result in this paper.

Some of the other results in Table 2 deserve comment. The point estimates of the intercept parameters for Options B, D, and E are based solely upon the choices by the roughly 325 members of the “stated preference” sub-sample (Sample 6). The estimates of WTP for the B, D, and E options suggest that people may be MOST willing to pay for the smallest program (Option B) and least willing to pay for (Option E), the most extensive program. This would appear to be the reverse of the usual "scope" effect, where people are expected to be willing to pay more if they get more.

This apparently anomalous result can be explored by appealing to other data we collected concerning individual preferences. All respondents were asked a preliminary question concerning "how interested" they were in the goal of replacing fossil fuel energy with renewable energy sources, and the goal of planting trees on public lands in upstate New York. The scale ranged from 1 (not at all interested) through 10 (very interested). About 9% of the sample failed to respond to each of these questions. Only 45% reported a rating of 6 or more for the renewable energy question (and only 17% rated it a 10); 70% reported a rating of 6 or more for the tree-planting question (and 39% rated it a 10). Respondents were clearly much more enthusiastic about the tree-planting exercise than about the renewable energy issue. It is possible that some respondents thought that the renewable energy would be provided free of charge to the “other households” mentioned in the survey.

Unlike the other sub-samples, additional information was gathered from the 325 respondents to the stated preference question involving all five program options. These individuals were asked to rate their satisfaction with each of the five options on a scale of 1 (not satisfied at all) to 10 (very satisfied) under the assumption that each could be implemented *at no cost to them*. The proportion of individuals recording a score of 6 or higher for each of the options is A:8%, B:24%, C:42%, D:66%, and E:78%. If the programs are free, more is clearly better. However, if the programs are costly, people would rather do something than nothing, but do not want to do anything more than the minimum. This certainly raises the issue of a "warm glow" effect.

The second horizontal panel of Table 2 gives the estimated ordered-logit threshold values, relevant for the multiple-bounded value information from Sample 5. Recall that an ordinary binary logit model is simply a special case of an ordered logit with only two categories of response. In the ordinary logit case, the threshold between the two categories (willing to pay, not willing to pay) is set arbitrarily to zero. The corresponding "zero-level" for the five-category multiple-bounded data should fall somewhere in the middle interval if data for this sample are being combined with data from the referendum sample (MDC). This middle interval is bounded by α_1 and α_2. We would therefore expect to see α_0 and α_1 negative, and α_2 and α_3 positive. When different dispersion parameters are allowed for each sample, this expectation appears to be borne out.

The third panel of Table 2 describes the estimated factors of proportionality in the error dispersions for each data type. These results are pertinent to the debate about the effects of survey instrument complexity on the cognitive difficulty experienced by respondents (DeShazo and Fermo, 1999). Survey Version 1 (the actual purchase decision sample, at $6) is defined as the numeraire sample, with dispersion parameter κ_1 normalized here to unity. Note that in order to ensure positive dispersions, we estimate the *logs* of the multiples of this dispersion factor for each sample. For ease of interpretation, Table 2 reports the corresponding *levels* of these estimated dispersion factors. (However, the associated t-test statistics still refer to the logged parameters.[14]) The asymptotic t-ratios can be used to test the hypothesis that the logarithm of the relevant parameter is zero (or, equivalently, that the estimated factor of proportionality for the dispersion is one, so that the dispersion is the same as for the numeraire sample).

The varying-bid dichotomous choice variant (Survey Version 2), displays an error dispersion parameter κ_2 that is about 1.7 times as large as the dispersion for Survey Version 1 (the actual choices). Contingent dichotomous choice valuation data appears to be noisier than data on actual choices.

Conventional wisdom among researchers who have worked extensively with open-ended and

---

[14] The conversion to levels accounts for the inconsistency in some of the signs on the parameter point estimates and the t-ratios.

referendum data holds that open-ended value estimates should be less variable. Here, however, the error dispersion in the open-ended data (Sample 3) is 2.6 times larger than for the numeraire 1-ACT sub-sample. The larger variance may be accounted-for by the existence of a handful of large outliers among the point values for willingness to pay in this sample. It may be that a Tobit-type conditional density based upon a logistic distribution is not fully compatible with the true distribution of WTP values in this sample. Alternately, these outliers may be just that--random, but influential, anomalies.[15]

For the payment card sample, however, the dispersion is *not* significantly different from that in the Version 1 (numeraire) sample. For these data, then, payment card elicitation seems to involve no more noise than do the actual purchase decisions. This result may or may not be generalizable. However, dispersion in the multiple-bounded (MB) data (Sample 5) is 3.5 times that in Version 1. This may belie greater cognitive challenges associated with this elicitation format.

The dispersion measures for the pairwise and joint stated preference samples are also larger than for the actual choice data, being 1.7 times as large in the pairwise case. This finding is satisfying in that one would expect this dispersion to be very much like that for the similar dichotomous-choice sample of Version 2. The joint stated preference sample, however, is 3.2 times as large as that for the actual choice data (and about twice that of the pairwise contingent choice data).[16]

The common indirect utility parameters estimated by the pooled-data model produce a single common point estimate of WTP for Option C, which is reported in the fourth panel of Table 2. The value for the model with differing dispersion parameters is $2.28, which is larger than the separate estimates for the open-ended and payment card methods, but less than the separate estimates for the

[15] In future research with these data, we plan to explore the prospect for more general distributions that allow for skewness. Needed is a distribution with the desirable properties of the logistic, but with at least one additional shape parameter. The logistic model conforms nicely with the concept of utility maximization, but there will be other, more general distributions that may prove useful.

[16] In this paper, we do not account for the simultaneity among the pairwise and joint choices for Survey Version 6, which are made by the same individuals.

referendum, multiple-bounded, and stated-preference samples.

For the fitted distributions of WTP for each sample, the expected values are the same but the dispersions are different. Different dispersion means that the proportion of the population that would be predicted to vote in favor of the policy (at any particular cost) will differ across elicitation methods. The last main panel of Table 2 shows the consequences of differing fitted dispersions for the predicted portion of the represented population that would be willing to pay at least $6 for Option 6. Despite identical *mean* WTP, these proportions are 23% for the actual (ACT) data, 33% for the varying referendum (MDC) data, 39% for the open-ended (OE) data,[17] 22% for the payment card (PC) data, 23% for the multiple-bounded (MB) data, 34% for the pairwise stated preference (SP_P) and 41% for the multiple-choice stated preference (SP_J) data.

6. Directions for Future Research

a.) Heterogeneous Preferences

Our estimating algorithms are set up to accommodate systematic varying parameters, rather than scalar parameters, for each of the five main indirect utility-difference parameters that are estimated in the models described above. Preliminary results indicate that with two shift variables on the intercept term and four on the slope term, it is possible to narrowly reject identical preference parameters under conventional likelihood ratio tests. However, due to the complexity of these models, we reserve an examination of these more-elaborate specifications for a subsequent paper.

[17] Note that a logistic distribution adapted to a Tobit-like specification may be inappropriate for these open-ended data. The outliers that produce a large estimated dispersion parameter also contribute to a large upper tail for the fitted distribution.

b.) Different Functional Forms for the Indirect Utility Function

In this paper, we have used a very simple functional form for the indirect utility-difference function that is linear in the price of the program. More exotic functional forms could certainly be explored, subject to the requirement that the corresponding formula for maximum willingness to pay must have an empirically tractable functional form. The more complicated the indirect utility-difference, the more unwieldy will be the associated WTP formula, and one must bear in mind that the error term from the former must also be carried through the derivation.

c.) Beginning with the Empirical Open-Ended Distribution

In this paper, we lean heavily on the tradition of using logit-type models to analyze discrete-choice contingent valuation data. This logit choice reflects the underlying extreme value distribution that is consistent with the utility theory underlying random utility models. However, the assumption of logistic errors derives only from this theory and represents a maintained hypothesis. In this paper, propagating the logistic errors through to the Tobit-type regression model used for the open-ended data maintains the utility-theoretic functional form and the stochastic structure throughout the model, but does not provide a particularly satisfying fit for the open-ended data.

It is possible that we could make better use of the open-ended data. Specifically, the open-ended sub-sample is the only one which directly gives us an array of point values for WTP. Some of these are large positive outliers, and these tend to drive up the implied error variance for this sample. It may be fruitful to "begin" with the open-ended data and to choose a functional form for the distribution of WTP values that is consistent with the observed marginal distribution of values for this sample. Perhaps one could then work backwards from this distribution to an alternative set of stochastic assumptions for the other elicitation modes.

d.) Explaining Error Dispersions

We are presently attempting to explain the observed differences in the magnitudes of the error dispersion parameters across elicitation modes. The present analysis shows that the implied error dispersions vary systematically with modes, and since modes were varied randomly across respondents, we do not expect the different modes to be picking up systematic differences in respondent characteristics across samples. However, even within a mode, there may be heteroscedasticity in WTP across individuals according to their characteristics. Cameron and Englin (1997) demonstrated heteroscedasticity in dichotomous choice contingent valuation data with respect to the amount of past experience each individual had with the environmental good in question. If the data-generating process for the errors (as a *function* of observable individual characteristics) differs systematically across samples, this could account, in part, for the different observed dispersions.

e.) More-General Specifications for the Multiple-Bounded Data

The α_j thresholds can be constrained to be identical for all 13 questions posed to each individual, or be allowed to differ by question by question. For each sample, then, there are potentially 39 additional threshold parameters to estimate, in addition to the basic parameters for the indirect utility function. However, since the first and 13th bid values were the same for all three versions of the multiple bounded instrument, it is not possible to estimate distinct sets of three thresholds for either question 1 or question 13. Questions 11 and 12 also pose problems for separate estimation since insufficient numbers of respondents chose the "Probably Yes" or "Definitely Yes" responses for these high bid values. Analysis is limited to a search for systematic tendencies in questions 2 through 10.

The estimated thresholds in an ordered logit model are typically interpreted as levels of the latent propensity variable, scaled by the error standard deviation of the conditional distribution of the propensity variable. Each of the three samples in this survey version may exhibit a different standard deviation for its error term. Likewise, we might allow for each question of the thirteen questions to exhibit a distinct error standard deviation. This lets us check for respondent fatigue (perhaps) if precision declines systematically with consecutive questions. Alternately, if precision increases with questions, perhaps learning (or crystallization of values) is occurring and respondents are increasingly able to identify their latent values as they progress through this list of choice questions.

7. Conclusions

For almost two decades, researchers have been puzzled by discrepancies among the values of non-market environmental goods according to the type of elicitation methods used to assess these values. Many past comparisons have been hampered by the need to use samples collected at different points in time, or from different populations, or via survey instruments that differ in other ways besides just the elicitation method employed. Early comparisons involved simply an inspection of the different point estimates of value produced by different methods, without conformable models or any opportunity for rigorous assessment of whether the values are statistically significantly different.

Our approach differs from earlier ones in a number of important ways. First, our different elicitation methods are part of a careful experimental design across split samples. All other aspects of the survey are carefully controlled so that any differences in the valuation results across samples can be attributed *only* to the different elicitation methods used.

Second, rather than comparing the mean WTP values implied by the different samples with their varying "treatments," we focus our efforts on the estimation of the underlying preference function parameters. This emphasis upon characterizing the preference function allows us to specify functional forms and stochastic structures that are entirely compatible across all of the different types of consumer choice scenarios that correspond to each elicitation method. The key insight is that if two or more elicitation methods imply the identical preference function for individuals in each sample, then these individuals will also have identical expected WTP. This is because the formula for expected WTP is derived from the indirect utility-difference function that is the foundation of all of our models.

Third, we zero in on the issue of heteroscedasticity across methods. In the context of dichotomous choice methods of elicitation, we do not estimate the indirect utility-function parameters individually. They are estimable only up to a scale factor (which is the dispersion parameter for the underlying latent utility-difference conditional distribution). Comparisons of the estimated parameters across different samples are invalid if the scale factors (the dispersion parameters) are different in the two groups. We allow the our utility-differences for each elicitation mode to exhibit whatever degree of dispersion the data dictate (subject to normalization on any one arbitrarily chosen sample).

What do we find? The evidence confirms the growing concensus that it is essential to accommodate heteroscedasticity across elicitation methods. With otherwise homogeneous preferences across individuals, there appears to be heteroscedasticity across elicitation methods, and this heteroscedasticity seems more-or-less consistent with the findings in the stated preference (conjoint analysis) literature that different methods lead to different “scale factors” (the common term for error dispersions). Even with the more than 7000 choices that we employ to fit our most elaborate specification that pools the data for all types of elicitation methods, we are not able to reject the hypothesis of identical indirect utility-difference functions across elicitation methods (i) provided preferences are assumed to be common across sociodemographic groups and (ii) provided we allow for heteroscedasticity across methods. If we attempt to impose homoscedasticity, however, the restriction of identical preference parameters is rejected.

On the topic of heteroscedasticity, there have been a couple of attempts to explain apparent differences in the amount of random noise associated with responses to different elicitation methods. Harrison (1989) and Smith and Walker (1993a,b) have suggested, bases on their laboratory results, that valuations will be more variable the lower the opportunity cost to respondents of deviating from the rational decision. Smith and Walker (1993b) show that bid function slopes are insensitive to these opportunity costs, but response variability can be decreased markedly by imposing larger opportunity costs.

It is indeed interesting to confirm that indirect utility differences appear to be heteroscedastic across elicitation methods and that ignoring this fact could contribute to apparent divergences between expected WTP estimates across methods. This cannot, however, be the end of the story. As researchers, we are now compelled to explain why error variances differ across methods. Pursuit of empirical models that can capture the sources of systematic variation in error variances across value elicitation methods is clearly high on the research agenda.

REFERENCES

Adamowicz, W.L., P. Boxall, M. Williams, and J. Louviere (1998) “Stated Preference Approaches for Measuring Passive Use Values: Choice Experiments and Contingent Valuation,” *American Journal of Agricultural Economics* 80(1) 64-75.

Adamowicz, W.L., J. Louviere, and M. Williams (1994) "Combining Stated and Revealed Preference Methods for Valuing Environmental Amenities," *Journal of Environmental Economics and Management* **26**, pp. 271-296.

Adamowicz, W.L., J. Swait, P.C. Boxall, J. Louviere, and M. Williams (1997) "Perceptions versus Objective Measures of Environmental Quality in Combined Revealed and Stated Preference Models of Environmental Valuation," *Journal of Environmental Economics and Management* **32**, pp. 65-84.

Boxall, Peter C., Jeffrey Englin, and Wiktor L. Adamowicz (1998) "Valuing Undiscovered Attributes: A Combined Revealed-Stated Preference Analysis of North American Aboriginal Artifacts," mimeo, Department of Applied Economics and Statistics, University of Nevada, Reno.

Boyle, Kevin J., and Richard C. Bishop (1988) "Welfare Measures Using Contingent Valuation: A Comparison of Techniques," *American Journal of Agricultural Economics* **70**(1), 20-28.

Boyle, Kevin J., F. Reed Johnson, Daniel W. McCollum, William H. Desvousges, Richard W. Dunford, and Sara P. Hudson (1996), "Valuing Public Goods: Discrete versus Continuous Contingent-Valuation Responses," *Land Economics* **72**(3), August, pp. 381-96.

Brown, Thomas C., Patricia A Champ, Richard C. Bishop, and Daniel W. McCollum (1996) "Which Response Format Reveals the Truth about Donations to a Public Good?" *Land Economics* **72**(2), May, pp. 152-66.

Barrett, C., T.H. Stevens, and C.E. Willis (1996) "A Comparison of CV and Conjoint Analysis in Groundwater Valuation," *W-133 Benefits and Costs Transfer in Natural Resource Planning*, 9, 79-104.

Cameron, Trudy Ann, and Jeffrey Englin (1997) "Respondent Experience and Contingent Valuation of Environmental Goods," *Journal of Environmental Economics and Management* **33**(3), 296-313.

Cameron, Trudy Ann and Daniel D. Huppert (1989) "OLS Versus ML Estimation of Non-Market Resource Values with payment Card Interval Data," *Journal of Environmental Economics and Management*, 17, 230-246.

Cameron, Trudy Ann, W. Douglass Shaw, and Shannon Ragland (1998) "Nonresponse Bias in Mail Survey Data: Salience vs. Endogenous Survey Complexity," in *Valuing the Environment Using Recreation Demand Models*, Joseph A. Herriges and Catherine L. Kling (eds.) Edward Elgar Publishing Ltd., forthcoming.

Cummings, Ronald G., et al. (1997), "Are Hypothetical Referenda Incentive Compatible?" *Journal of Political Economy* 105, 609-621.

DeShazo, George M., and German D. Fermo (1999) “Understanding the Effects of Choice Set Complexity on Consumer’s Choice Behavior: An Application to Recreational Demand,” paper presented at the W-133 Annual Meeting, Tucson, AZ, February 24-26, 1999.

Desvousges et al. (1993) "Measuring Natural Resource Damages with Contingent Valuation: Tests of Validity and Reliability," in Hausman, Jerry A. (ed.) *Contingent Valuation: A Critical Assessment*, New York: Elsevier Science, 91-159.

Ethier, Robert, J. Clark, Greg Poe and William D. Schulze (1997) “A comparison of hypothetical phone and mail contingent valuation responses with actual willingness to contribute to green pricing electricity programs,” **American Journal of Agricultural Economics 79**(5) 1719.

Halvorsen, Bente, and Kjartan Soelensminde (1998) “Differences between Willingness-to-Pay Estimates from Open-Ended and Discrete-Choice Contingent Valuation Methods: The Effects of Heteroscedasticity,” *Land Economics* **74**(2), 262-82.

Harrison, Glenn (1989) "Theory and Misbehavior in First Price Auctions," *American Economic Review*, 749-62.

Hensher, David, Jordan Louviere, and Joffre Swait (1999) “Combining Sources of Preference Data,” *Journal of Econometrics* **89**(1-2), 197-221.

Holmes, Thomas P., and Randall A. Kramer (1995) "An Independent Sample Test of Yea-Saying and Starting Point Bias in Dichotomous-Choice Contingent Valuation," *Journal of Environmental Economics and Management* **29**, 121-132.

Huang, Ju-Chin, and V. Kerry Smith (1998) “Monte Carlo Benchmarks for Discrete Response Valuation Methods,” *Land Economics* **74**(2) 186-202.

Johnson, Rebecca L., N. Stewart Bregenzer, and Bo Shelby (1990) "Contingent Valuation Question Formats: Dichotomous Choice versus Open-Ended Responses," in Johnson, Rebecca L, and Gary V. Johnson (eds.) *Economic Valuation of Natural Resources: Issues, Theory, and Applications*. Social Behavior and Natural Resources Series, Boulder and Oxford: Westview Press, 193-203.

Kealy, Mary Jo, and Robert W. Turner (1993) "A Test of the Equality of Closed-Ended and Open-Ended Contingent Valuations," *American Journal of Agricultural Economics*, **75**(2) May, 321-31.

Kahneman, D. and A. Tversky (1984) "Choices, Values and Frames," *American Psychologist* **39**, 341-50.

Kramer, Randall A., and D. Evan Mercer (1997) “Valuing a Global Environmental Good: U.S. Residents’ Willingness to Pay to Protect Tropical Rain Forests.” Land Economics **73**, 196-210.

Loomis, John, Thomas Brown, Beatrice Lucero, and George Peterson, "Evaluating the Validity of the Dichotomous Choice Question Format in Contingent Valuation," *Environmental and Resource Economics*; **10**(2), September 1997, 109-23.

Loomis, John B., Michael Lockwood, and Terry DeLacy (1993) "Some Empirical Evidence on

Embedding Effects in Contingent Valuation of Forest Protection," *Journal of Environmental Economics and Management*, **25**(1), Part 1, July, 45-55.

Loomis, John B. (1990) "Comparative Reliability of the Dichotomous Choice and Open-Ended Contingent Valuation Techniques," Journal of Environmental Economics and Management, **18**(1), January, 78-85.

Mackenzie, John, "A Comparison of Contingent Preference Models," *American Journal of Agricultural Economics*; **75**(3), August 1993, 593-603.

McFadden, Daniel (1994) "Contingent Valuation and Social Choice," *American Journal of Agricultural Economics*, **76**(4:Nov), 689-708.

Milon, J. Walter (1989) "Contingent Valuation Experiments for Strategic Behavior," Journal of Environmental Economics and Management, 17(3), November, 293-308.

Welsh, Michael and Greg Poe (1998) "Elicitation effects in contingent valuation: Comparisons to a multiple bounded discrete choice approach" *Journal of Environmental Economics and Management* **36** (2) 170-185.

Ready, Richard C., J.C. Buzby, and D. Hu (1996) "Differences Between Continuous and Discrete Contingent Valuation Estimates," *Land Economics*, **72**(3:Aug), 397-411.

Schulze, William D., G. McClelland, D. Waldman, and J. Lazo (1996) "Sources of Bias in Contingent Valuation," in D.J. Bjornstad and J.R. Khan, eds., *The Contingent Valuation of Environmental Resources*, Edward Elgar Publishers, Cheltenham, UK.

Sellar, Christine, J.R. Stoll, and J.-P. Chavas (1985) "Validation of Empirical Measures of Welfare Change: A Comparison of Non-Market Techniques," *Land Economics* **40**(2:May), 156-175.

Smith, Vernon L., and J.M. Walker (1993a) "Rewards, Experience, and Decision Costs in First Price Auctions," *Economic Inquiry* **31**, 237-245.

Smith, Vernon L., and J.M. Walker (1993b) "Monetary Rewards and Decision Cost in Experimental Economics," *Economic Inquiry* **31**, 245-261.

Stevens, Thomas H.; Barrett, Christopher; Willis, Cleve E. (1997) "Conjoint Analysis of Groundwater Protection Programs," *Agricultural and Resource Economics Review* **26**(2), October, 229-36.

Tversky, A., S. Sattath and P. Slovic (1988) "Contingent Weighting in Judgment and Choice," *Psychological Review* **95**, 371-384.

Table 1
Pairwise Comparisons of WTP Estimates by Different Elicitation Methods

	Solo	2-MDC		3-OE		4-PC		5-MB		6-SP	
		same κ	diff. κ	same κ	diff. κ	same κ	diff. κ	same κ	diff. κ	same κ	diff. κ
1-ACT	-62.940 $ (n/a) np=1 n=116	**-520.614** $2.57 np=2 n=851	*-517.559* $2.84 np=3 κ_2=1.91*	**-772.424** $0.29 np=2 n=356	*-769.134* $1.31 np=3 κ_3=2.06	*-746.834* $2.15 np=2 n=387	*-746.718* $2.09 np=3 κ_4=0.91	**-621.291** $0.00 np=6 n=4956	*-617.165* $0.31 np=7 κ_5=2.30	*-315.947* $1.75 np=6 n=1373	*-313.953* $3.45 np=7 κ_{6P}=2.46 κ_{6J}=4.49*
2-MDC	-454.619 $2.84 np=2 n=735			**-1165.927** $2.01 np=2 n=975	*-1162.021* $2.54 np=3 κ_3=1.53*	**-1151.187** $2.67 np=2 n=1006	*-1139.362* $2.33 np=3 κ_4=0.55*	**-1017.392** $2.07 np=6 n=5575	*-1008.844* $2.84 np=7 κ_5=2.16*	*-705.746* $2.89 np=6 n=1992 κ_{6J}=1.89*	*-705.728* $2.90 κ_{6P}=1.05 κ_{6J}=1.94*
3-OE	-706.194 $1.30 np=2 n=240					**-1440.185** $1.85 np=2 n=511	*-1390.320* $2.01 np=3 κ_4=0.37*	**-1262.349** $1.11 np=6 n=5080	*-1260.419* $1.31 np=7 κ_5=1.35*	*-959.207* $1.76 np=6 n=1497 κ_{6J}=1.56	*-958.089* $1.93 np=7 κ_{6P}=0.68 κ_{6J}=1.28
4-PC	-683.778 $2.09 np=2 n=271							**-1287.151** $1.80 np=6 n=5111	*-1238.003* $2.09 np=7 κ_5=3.69*	**-939.761** $2.23 np=6 n=1528 κ_{6J}=	*-935.281* $2.16 np=7 κ_{6P}=1.84* κ_{6J}=3.43*
5-MB	-554.225 $3.08 np=6 n=4840									**-808.499** $3.78 np=10 n=6097	*-805.236* $3.45 np=11 κ_{6P}=0.48* κ_{6J}=0.88
6-SP	-251.012 $3.45 np=6 n=1257 κ_{6J}=1.83*										

KEY: Cells for "same κ" models contain maximized value of the log-likelihood function, fitted WTP for option C, number of estimated parameters, number of choices employed in model; Cells for "diff. κ" models contain maximized value of the log-likelihood, fitted WTP for option C, number of estimated parameters, point estimate of κ_j factor for sample j. Note that 6-SP data allows differing κ_{6J} for multiple choice (versus pairwise choices) in all models. Bold-face log-likelihood indicates restrictions (against separate models) are rejected at 5% level of significance; italics indicates failure to reject.

Table 2: Pooled Samples; Utility Parameters Constrained

Type of Parameter	Common κ across data types	Differing κs across data types
Utility parameters:		
$\exp(\beta_{11})$ * Option C_i +	-1.048 (-7.45)**	-0.310 (-1.40)
$\exp(\beta_{21})$ * Option B_i +	0.152 (0.69)	0.724 (2.32)**
$\exp(\beta_{31})$ * Option D_i +	-2.431 (-0.79)	-1.397 (-0.64)
$\exp(\beta_{41})$ * Option E_i -	-16.26 (-0.10)	-19.86 (-0.07)
$\exp(\beta_{51})$ * bid_i	-1.756 (-42.42)**	-1.136 (-68.88)**
Ordered logit thresholds:		
α_0	-1.228 (-11.16)**	-3.285 (-4.61)**
α_1	-0.6542 (-7.48)**	-1.481 (-3.37)**
α_2	-0.1525 (-2.39)**	0.09364 (0.25)
α_3	0.7542 (7.09)**	3.047 (3.72)**
κ multiples:[a]		
κ_1 (actual)	1.0	1.0
κ_2 (MDC)	1.0	1.732 (2.85)**
κ_3 (OE)	1.0	2.575 (5.22)**
κ_4 (PC)	1.0	0.9505 (-0.30)
κ_5 (MB)	1.0	3.514 (5.83)**
κ_6 ($SP_{pairwise}$)	1.0	1.735 (2.01)**
κ_7 (SP_{joint})	1.813 (2.16)**	3.236 (3.54)**
Fitted WTP at means of data		
Option C	$2.03	$2.28
Proportion WTP $6		
1 - ACT	33.5 %	23.3 %
2 - MDC	"	33.4 %
3 - OE	"	38.6 %
4 - PC	"	22.2 %
5 - MB	"	23.3 %
6 - $SP_{pairwise}$	"	33.5 %
7 - SP_{joint}	40.7 %	40.9 %
Number of choices employed	7459	7459
Maximized Log L	-2784.699 [c]	- 2714.733 [d]

[a] κ multiples are estimated as powers of e to ensure that they remain positive. For ease of interpretation, the point estimates have been exponentiated. The t-test statistics, however, remain tests of the hypotheses that the estimated exponents are zero. If the exponent is zero, the κ multiple is unity (exp(0)) and the dispersion for the sample in question is not different from that of the numeraire 1-ACT sample.

[b] Percentages differ only because conditional dispersion of WTP distribution differs across data types. E[WTP] is constrained to be identical for all samples. Logistic distributional assumption is strong (especially in 3 - OE Tobit-type specification).

[c] The sum of the maximized log-likelihood values for the separately estimated models (see Table 1) is -2712.768. There are 25 implicit parameters across the six differ data types, although only 21 of them can be identified. There are only ten parameters in this pooled-data model. The likelihood ratio test statistic for the restrictions embodied in this model is 143.8. Regardless of how one counts the number of restrictions they are clearly rejected.

[d] There are five fewer restrictions in this model, since the κ parameters for each data type are allowed to differ. The likelihood ratio test statistic for the restrictions in this model, compared to the separately estimated models for each data type, is only 3.93. Thus, we cannot reject identical indirect utility-difference parameters, providing the κ parameters are allowed to differ.

Appendix

Non-response Analysis

In any survey-based research, it is important to consider the potential effects of unit non-response on the statistical relationships revealed from estimates based solely on respondents. Cameron, Shaw, and Ragland (1999) discusses the competing influences of survey topic salience and endogenous survey complexity in determining the probability that a given randomly selected mail survey recipient will elect to complete and return the survey instrument. Since the present study concerns passive use values, rather than active use values, it is spared most of the ills of endogenous complexity. In this case, the respondent's level of activity involving the goods to be valued does not affect the amount of time and/or energy required to fill out the questionnaire.

Fortunately, the type of elicitation method on any given individual's survey is completely unrelated to their individual attributes, since the elicitations are randomly assigned across participants. Thus the type of elicitation method cannot be correlated with any factor that might influence propensity to complete and return a questionnaire. Nevertheless, we retained the five-digit zip codes associated with every questionnaire mailed out. We matched each targeted household with the 1990 Census data for its zip code in an attempt to find any systematic differences in response propensities.

It is possible that respondents may be systematically less likely to cooperate with some types of surveys in our study than with others. If the observed nonresponse/response (a 0,1 variable) is regressed using a probit model on a set of dummy variables for elicitation mode (using the 2-MDC mode as the numeraire), we find no statistically significant differences in response rates across elicitation methods.

But we have also explored the effects of Census zip code demographics on response propensities and have found that a wide variety of zip code level variables, used individually, make a statistically significant difference to response rates. PPUBINC (proportion on public assistance income) has a negative effect on response probability; IP200UP (proportion with income greater than 200% of the poverty level) has a positive effect on response probability; POWNOCC (proportion owner-occupied housing) has a positive effect on response probability, as does MEDGRRNT (median gross rent). PWHITE (proportion white) has a positive effect, and PBLACK (proportion black) has a negative effect. LOWED (proportion with low education levels) has a negative effect, and LANGIS (proportion language-isolated) also has a negative effect on response probabilities. Due to the high degree of collinearity among these variables, a model that includes all of them in the same specification does not reveal any individually statistically significant coefficients. The likelihood ratio test statistic for the hypothesis that all slope coefficients are jointly zero, however, is 115.8, which handily exceeds the critical value for a $\chi^2(8)$ distribution.

The relevant issue for comparisons of implied WTP values across elicitation modes, however, is whether there are systematic differences in response rates as a function of elicitation modes. Thus we have included not only dummy variables for elicitation modes and Census zip code demographics, but also a host of interaction terms between the two categories of variables. The relevant hypothesis test is a likelihood ratio test for this unrestricted model compared to a restricted model with no elicitation mode dummies or interaction terms. Results show that the maximized value of the probit log-likelihood is minimally compromised by restricting the elicitation-mode effects to zero, so we can safely argue that there are unlikely to be any quantitatively significant variations in response rates across elicitation modes.

Response rates do not depend on elicitation methods, but they do differ according to demographic characteristics of the neighborhood (zip code) to which the survey was sent. Thus, in any model designed to estimate the overall social value of the GreenChoice™ program, it will be important to compensate for these differing response propensities.

Table A.1

Response Rates for Survey Variants

Survey Type	Phone 1-DC $6	Mail 2-DC $.50,1,2,4 ,6,9,12	Mail 3-OE	Mail 4-PC	Mail 5-MB	Mail 6-SP
Intended Sample	250	1400	500	500	900	600
Adjusted Sample (undeliv. no phone not NMPC customer)	199	1242	436	435	781	525
Completes	145	831	278	298	522	345
Response Rate (%)	70.4	66.9	63.8	68.5	66.8	65.7

ATTACHMENT TO FINAL REPORT

National Center for Environmental Research and Quality Assurance
NCERQA Grant Final Report
Executive Summary (3-5 pages)

Date of Report: March 31, 1999

EPA Grant Number: R824688

Title: Can Contingent Valuation Measure Passive-Use Values?

Investigators: Trudy Ann Cameron, Jeremy E. Clark, Robert G. Ethier, Gregory L. Poe, Daniel Rondeau, Steven K. Rose, and William D. Schulze

Institution: Cornell University

Research Category:

Project Period: 10/1/95-12/31/98

Description and Objective of Research:

This project attempts to compare and validate alternative elicitation methods for contingent valuation. The commodity used to test alternative methods was based on an actual green choice offering made by the Niagara Mohawk Power Company to fund non-renewable generation for 1200 homes and to plant 50,000 trees. Given the cooperation of the Niagara Mohawk Power Company, we were able to compare actual participation in the program solicited through a telephone interview with hypothetical participation as estimated from the various elicitation procedures employing both telephone and mail interviews.

Five papers resulted from the research project. Each paper represents a single aspect of the research and may be read independently by readers interested in the specific issues covered.

Summary Findings and Conclusions:

Daniel Rondeau, William D. Schulze and Gregory L. Poe, Voluntary Revelation of the Demand for Public Goods Using a Provision Point Mechanism, (Forthcoming in *Journal of Public Economics*).

A one-shot provision point mechanism with money-back guarantee and proportional rebate of excess contributions is tested in an induced value framework and in experimental environments chosen to mimic field conditions. The results show that this relatively simple mechanism is empirically demand revealing in the aggregate when used with large groups of students who have heterogeneous valuations for the public good. Approximately demand revealing behavior was obtained under three alternative information conditions. These results are an important step in the design of a mechanism simple enough to allow field applications, but capable of efficiently providing public goods through voluntary contributions.

Steven K. Rose, Jeremy Clark, Gregory L. Poe, Daniel Rondeau, William D. Schulze, The Private Provision of Public Goods: Tests of a Provision Point Mechanism for Funding Green Power Programs, Cornell University Environmental and Research Economics Paper Series, 97-02, 1997.

This paper utilizes field and laboratory experiments to test the use of a provision point mechanism to finance renewable energy programs, commonly known as green pricing programs. In contrast to most green pricing programs, relatively high participation is found in the field, while laboratory results suggest that demand revelation is achieved by the mechanism in a single shot environment with a large group of potential participants.

Gregory L. Poe, Jeremy E. Clark and William D. Schulze, Can Hypothetical Questions Predict Actual Participation in Public Programs? A Contingent Valuation Validity Test Using a Provision Point Mechanism, (*Journal of Environmental Economics and Management* – revise and resubmit status).

Past field validity tests of contingent valuation have relied on voluntary contribution mechanisms to elicit actual willingness to pay, and are thus likely to overestimate hypothetical bias because of free riding in the actual contributions. This chapter argues that provision point mechanisms -- which have recently been shown to approximately reveal demand in large group public goods experiments -- should instead be used in contingent valuation validity testing, and employs such a mechanism in a validity study of green electricity pricing. Some upward hypothetical bias is found even when this improved mechanism is used. Calibration of hypothetical responses is also explored.

Robert G. Ethier, Gregory L. Poe, William D. Schulze, and Jeremy Clark, A Comparison of Hypothetical Phone and Mail Contingent Valuation Responses for Green Pricing Electricity Programs, (Forthcoming in *Land Economics*).

This study provides the first contingent valuation phone-mail comparison that meets current standards for response rates, draws from a general population, is relevant to the valuation of general environmental goods, and allows comparisons with actual participation rates. Social desirability effects are found to be more prevalent in phone responses to subjective questions, but do not appear to affect hypothetical participation decisions: calibrated and uncalibrated hypothetical participation rates are statistically similar across modes. As such, neither mode appears to dominate from the perspective of providing more valid estimates of actual participation decisions.

Trudy Ann Cameron, Gregory L. Poe, Robert G. Ethier, and William D. Schulze, Alternative Non-market Value-Elicitation Methods: Are the Underlying Preferences the Same?, (Submitted to *Journal of Environmental Economics and Management*).

In a unique survey, six different random sub-samples of respondents were presented with an opportunity to value the identical environmental good, each via a different elicitation method. The methods include: an actual purchase decision at a single price, a referendum format with differing prices, an open-ended format, a payment card format, a multiple-bounded format, and a stated choice among an extended set of five alternatives (including the basic good plus three additional alternatives). We employ a common underlying indirect utility function and a stochastic structure that is also fully compatible across methods, allowing us to pool all of these different types of valuation data in one unified model. The different types of valuation data are entirely compatible with the same underlying set of homogeneous preferences, providing heteroscedastic errors across elicitation methods are permitted.

Supplemental Keywords: public goods, voluntary contributions, provision point, experiments, information group size, green pricing, renewable energy, free riding, altruism,

Relevant Web Sites:

***NCERQA will not edit without consultation with Principal Investigator.**